DOCUMENT RESUME

ED 360 833 FL 021 410

AUTHOR	Davies, Alan
TITLE	Edinburgh Working Papers in Applied Linguistics, Number 4.
INSTITUTION	Edinburgh Univ. (Scotland).
REPORT NO	ISSN-0959-2253
PUB DATE	93
NOTE	143p.; For individual papers, see FL 021 411-420. For earlier volumes, see ED 353 834 and ED 353 780.
AVAILABLE FROM	Edinburgh Working Paper in Applied Linguistics, Institute for Applied Language Studies, 21 Hill Place, Edinburgh EH8 9DP, Scotland, United Kingdom (Individual: surface mail, 5 British pounds and air mail, 7 British pounds; Institutional: surface mail, 6 British pounds, and air mail, 8 British pounds).
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Edinburgh Working Papers in Applied Linguistics; v4 1993
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Applied Linguistics; Case Studies; Ethnography; Foreign Countries; Higher Education; Instructional Innovation; Interpersonal Communication; Interviews; *Language Tests; Lecture Method; Native Speakers; Reading Comprehension; *Second Language Instruction; Second Languages; Tests; Transfer of Training; Translation; Writing (Composition)

ABSTRACT

"Edinburgh Working Papers" is intended to show a yearly cross-section of current work in Edinburgh's Department of Applied Linguistics and to elicit reactions and criticism. Articles in this volume include the following: "MSC Common Room Conversations: Topics and Terms" (Joan Cutting); "Speculation and Empiricism in Applied Linguistics" (Alan Davies); "Implementing Innovation in Language Education" (Gibson Ferguson); "Functional Controlled Writing" (Ardeshir Geranpayeh); "Development and Validation of a Translation Test" (Behzad Ghonsooly); "The Design and Validation of a Multi-Level Reading Comprehension Test" (Aileen Irvine); "Questions in Lectures: Opportunities or Obstacles?" (Tony Lynch); "Can Applied Linguists Do Ethnographic Interviews?" (Brian Parkinson); and "Influence of Languages Other than the L1 on a Foreign Language: A Case of Transfer from L2 to L3" (Matutin Sikogukira). (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 360 833

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

THE UNIVERSITY OF EDINBURGH

Number 4
1993

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Brian
Parkinson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL021410

BEST COPY AVAILABLE

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS

Number 4

1993

(ISSN 0959-2253)

Published jointly by the Department of Applied Linguistics
and the Institute for Applied Language Studies
University of Edinburgh
Scotland

General editor: Alan Davies
Assistant general editor and Issue 4 editor: Brian Parkinson

Subscription information

Edinburgh Working Papers in Applied Linguistics (EWPAL) is published in the third term of the University of Edinburgh year (April-June). Details of subscription rates and journal exchange arrangements are at the end of this issue.

Copyright

Copyright rests with the authors of the papers in this volume, who are free to submit their work for publication elsewhere, either as published here or in a revised version.

Contents

MSC COMMON ROOM CONVERSATIONS: TOPICS AND TERMS
Joan Cutting 1

SPECULATION AND EMPIRICISM IN APPLIED LINGUISTICS
Alan Davies 14

SEASPEAK
Alan Davies 26

IMPLEMENTING INNOVATION IN LANGUAGE EDUCATION
Gibson Ferguson 27

FUNCTIONAL CONTROLLED WRITING
Ardeshir Geranpayeh 40

DEVELOPMENT AND VALIDATION OF A TRANSLATION TEST
Behzad Ghonsooly 54

THE SIGNIFICANCE OF 'SIGNIFICANCE'
Martin Gill 63

THE DESIGN AND VALIDATION OF A MULTI-LEVEL READING COMPREHENSION TEST
Aileen Irvine 81

QUESTIONS IN LECTURES: OPPORTUNITIES OR OBSTACLES?
Tony Lynch 87

CAN APPLIED LINGUISTS DO ETHNOGRAPHIC INTERVIEWS?
Brian Parkinson 96

INFLUENCE OF LANGUAGES OTHER THAN THE L1 ON A FOREIGN LANGUAGE: A CASE OF TRANSER FROM L2 TO L3
Matutin Sikogukira 110

Preface

EWPAL provides an annual update on some of the work currently being carried out in applied linguistics by students and staff at the University of Edinburgh. The work of six authors from the Department of Applied Linguistics (DAL) and of one from the Institute for Applied Language Studies (IALS) is featured in **EWPAL 4** for the first time, and their contributions are particularly welcome.

The editorship of this issue requires some clarification. For issues 4 to 6 the editorship of EWPAL passes from IALS to DAL, and Alan Davies is the overall editor for these issues, Brian Parkinson the assistant editor. Issue 4 has, however, gone to press at a time when Alan is on sabbatical, and the assistant editor must therefore take responsibility for the final form of this issue. I have tried to edit with a light touch, removing as far as possible obvious errors and inconsistencies within and between articles, but I have not changed various forms such as generic masculine pronouns which I would not use myself but which still seem to be within the realm of choice.

I would like to acknowledge the willing help of the members of the Editorial Board - Esther Daborn, Cathy Benson, Alan Davies, Eric Glendinning, Joan Maclean, Liam Rodger and Sonia S'hiri - who made time to read and comment on the manuscripts submitted for **EWPAL 4**.

Thanks also go to my IALS colleague Elaine Bell for her efficient and invaluable assistance in turning contributors' 'final' versions into these published papers. The *final* final form of **EWPAL 4** also owes much to the expertise of Ray Harris and his colleagues at the Reprographics Department of the University of Edinburgh.

Brian Parkinson
Editor

April 1993

MSC COMMON ROOM CONVERSATIONS: TOPICS AND TERMS

Joan Cutting (DAL)

Abstract

This paper explores the difference between conversations of new acquaintances and those of established friends. Psycho- and sociolinguistic literature on the subject lacks a systematic grammatical and lexical approach to the analysis of distinguishing features. This paper describes a longitudinal study of the 1991-92 Edinburgh University Applied Linguistics MSc students' common room conversations investigating how the language of this discourse community changes over the duration of the course as the students' shared knowledge increases. The contextualisation cues to be examined in the full study are special terms and names, general nouns and verbs, exophoric reference, substitution and ellipsis. So far, only special terms have been analysed, and this paper discusses the trends revealed, pointing to areas requiring further exploration.

1. Introduction

1.1 Overall aim of the full study.

This paper describes the first step in a larger study that aims to analyse the way that the 1991-92 Edinburgh University Applied Linguistics MSc students' language was affected by their interacting over time through the MSc course. In the larger study I will concentrate on the effect of growing and changing knowledge areas on the language of the students, taking a pragmatic approach to the process of linguistic change. I have chosen to focus on certain grammatical and lexical features that depend for their meaning on knowledge of the situational context of the MSc course and interactions within its duration. I hypothesise that certain 'contextualisation cues,' to use Gumperz's (1982) term referring to linguistic features that contribute to the 'signalling of contextual presuppositions' (p.71), increase over time. These can be categorised as exophoric reference, substitution and ellipsis, special terms and names, and general nouns and verbs. My aim is to take a developmental view of the special language that evolves in this closed network group, in order to determine if the cues emerge over time in any particular order and how they relate to each other. I hope to show how the in-group's language could become increasingly inaccessible to an outsider to this MSc group as the number of implicit references to assumed knowledge areas grows. I aim to make a generalisable statement about the pragmatic nature and function of language of any group starting as strangers and becoming a discourse community, united by a common goal and interacting over an extended but defined period of time.

The first part of the present paper will begin by reviewing the literature on the language of social groups and the indicators of intimacy. Then individual hypotheses will be stated for each of the contextualisation cues to be examined in the larger study, and general hypotheses will be put forward to suggest how these cues might interrelate.

1.2 Specific aim of this step in the study.

The second half of the present paper will discuss results of analysis of special terms and other course-related terms used by the MSc students from the beginning to the end of the course.

2. Review of the literature

No study has followed through the interactions of a group of people as they become a discourse community, defined by its members' common goals, intercommunication mechanisms, particular genres and specific lexis (Swales 1990), to discover exactly how and when grammatical and lexical reference to shared knowledge develops over time as the assumed knowledge grows. Swales describes the academic class as forming a discourse community, but does not analyse how it forms:

> Somewhere down the line, broad agreement on goals will be established, a full range of participatory mechanisms will be created, information exchange and feedback will flourish by peer-review and instructor commentary, understanding the rationale of and facility with appropriate genres will develop, control of technical vocabulary in both oral and written contexts will emerge, and a level of expertise that permits critical thinking be made manifest.
>
> (Swales 1990:32)

Our MSc students fulfil Swales' criteria for a discourse community. They have the broadly agreed common public goal of passing the course; their mechanisms for communication are mainly face-to-face interaction, whether in tutorials or in the common room; these mechanisms they use to provide feedback, but also solidarity and relief from anxieties; they acquire a special lexis; they possess more than one genre (common room casual conversation is but one, with its two registers: course-related topics and non-course-related topics).

Some studies have described the language of social groups but they lack a suggestion of how exactly language changes to become the language of the social group. One of the best descriptions is that of Bernstein (1971): he lists the characteristics of restricted code, such as restricted lexical and syntactical alternatives, few subordinate clauses, metaphor, and says that they 'interact cumulatively and developmentally reinforce each other and so the effect of any one depends on the presence of the others' (Bernstein 1971:43), but gives no suggestion as to how this might happen. Our MSc students develop a restricted code in the sense that it is context-dependent and contains unspoken assumptions, just like that of the university students in Levy's (1979) study who talked about their selected course subjects in such a way that even the staff found it difficult to understand. The restricted code of our students in the Applied Linguistics common room contains elements of Bernstein's elaborated code: because of their course experience, their language, in particular their lexis, can be rational and abstract.

Those sociolinguists and psycholinguists who have considered how assumed knowledge areas and language change over time interdependently, refer to the change in superficial terms. Gumperz (1982) is one whose work is especially relevant to my study. He acknowledges that 'exclusive interaction with individuals of similar background leads to reliance on unverbalised and context-bound presuppositions in communication' (p.131), and lists contexualisation cues such as prosodic features, formulaic expressions, sequencing strategies and lexis and syntax. Unfortunately, he does not explore the area of lexis, syntax and phonology in depth. Kreckel (1981) recognises the dynamic nature of in-group language formation yet makes no attempt to consider the process whereby the in-group language might be formed. She describes the language of university students in terms of product alone: it consists of 'a multitude of in-group codes, discipline specific and social group specific...taking discipline or group-specific knowledge for granted' (Kreckel 1981:36).

Tannen (1984, 1989), in describing the high involvement style of those who regularly interact, mentions interpersonal involvement signals such as playful routines, irony, allusion, reference to familiar jokes and assumptions, ellipsis, indirectness, tropes, and imagery, yet she does not examine the order of emergence of these signals over time or the relation between each. Thus it is a static rather than a dynamic description: there is no suggestion of the routes from low involvement to high involvement; of how to get from one stage to another.

The present longitudinal study was undertaken to provide a systematic model for describing and hopefully predicting the process of language changes over time as individuals form a discourse community.

3. The study of cues: the full study

3.1 Hypotheses.

It is not only the background knowledge that can make a closed social network's conversations exclusive to an outsider to the group, but also the fact that the group members refer to that situational context in a particular way using contextualisation cues of reference, substitution and ellipsis, special terms and names and general words. Although the model of the cues in Figure (1) is my own, most of the classifications are Halliday's (1976). I hypothesise that as shared knowledge grows, the intertextual frequency and textual density of contextualisation cues increases and that the language of in-group members has more contextualisation cues than that of strangers.

Fig. (1) Contextualisation cues: indicators of in-group membership.

GRAMMAR

- Exophoric reference
  - personal
  - demonstrative
  - comparative
- Exophoric substitution and ellipsis
  - nominal
  - verbal (lexical)
  - unfinished sentences

LEXIS

- Special terms
- Proper nouns/names of people
- General words
  - nouns
  - verbs

The shared knowledge I see as falling into five main areas which I classify as K1 to K5. K1 is general knowledge of the world, including Edinburgh and the University; K2 is general knowledge of linguistics and language teaching (fields, notions, names, etc.); K3 is knowledge of DAL and IALS organisation typical of any MSc year (courses, programme deadlines, projects, classes, staff); K4 is knowledge of this particular MSc year (specific tasks, specific study groups, particular books and articles, special ways of referring to courses, students); K5 is shared knowledge of personal details of the interlocutor, the interpersonal context (interlocutor's family and origin, characteristics and interests).

Topics in these five areas can be grouped in two macro-categories: course-related and non-course-related knowledge. K2 to K4 always contain course-related topics (henceforth 'c topics') and K5 always contains non-course-related topics ('n-c topics'). K1 is generally non-course-related, although K1 c topics (K1(C)) are ones related to the context of the course, such as how to run a computer programme that fulfills the needs of a certain project and how to go about converting to an M.Litt. or applying for an IALS scholarship.

I hypothesise that c topics will be more impenetrable to an outsider than n-c topics. I predict that with time, (K2 to K5) c topics will be more frequent than (K1) n-c topics, and that this will cause the conversations to have larger impenetrable sections because of the co-existence of not only the assumed knowledge area but also the greater number of occurrences of contextualisation cues.

I now state the individual hypotheses for each contextualisation cue. The first set of contextualisation cues is lexical: I hypothesise that there will be an increase in special terms (technical and course-related) as shared knowledge grows; that the percentage of special terms out of all nouns in K2-K4 will increase over time (see Section 4). Paradoxically, I also predict an increase in what I call general course-related terms, by which I mean count nouns usually with zero article whose precise meaning is not clear since they are the first noun of a two- or three-word phrasal expression (Huddleston 1988:103) whose second/third word(s) (often a superordinate) is/are omitted. Thus their meaning varies from context to context.

e.g. OCT 25: 'Has anybody done their syntax?' (syntax tutorial task)

JAN 13: 'Are you going to stylistics?' (the stylistics lecture)

On the boundary between special terms and general words are superordinates. I predict an increase in superordinate course-related terms, substituting K2-K4 special terms, whose specific reference could be supplied by a pre-head modifier.

e.g. JAN 13: 'I'm not going to read the book again.' (the phonetics set book)

JAN 20: 'And the paper's due in next Friday.' (the core paper: the first project)

The second type of lexical contextualisation cue is that of proper nouns and names of people. Again, I hypothesise that there will also be an increase in a category of a general use of names of people which refer elliptically to something other than the people named, again as the first noun of a phrasal expression.

e.g. JAN 13: 'I haven't done any Chomsky.' (Chomsky revision or revision of materials about/by Chomsky)

Thirdly, I shall examine the lexical contextualisation cue of exophoric general words. I hypothesise that course-related words will be substituted increasingly by general words as K2-K4 grow.

e.g. JAN 13: 'I've done all the people.' (studied, thinkers)

Moving on to grammar, the fourth set of contextualisation cues to be analysed is exophoric reference. I predict an increase in the percentage of exophoric third person singular/plural existential pronouns and possessives out of all third person personals and an increase in demonstrative pronouns referring to course-related referents.

e.g. JAN 20: 'I mean you know what she's like. She's really fanatic.'

JAN 27: 'So I typed that thing up again after you'd gone.'

I predict an increase in the percentage of definite noun phrases on first mention: 'the' (non-generic) with special terms (K2-K4) and general nouns, out of all noun phrases.

NOV 7: 'So you've got the whole damn thing to do.'

I predict an increase in the percentage of exophoric comparative reference out of all comparative reference.

e.g. JAN 13: 'I feel more comfortable with the data stuff.'

The fifth set of contextualisation cues to be studied is exophoric substitution and ellipsis.

e.g. JAN 13: 'Which ones are you concentrating on?'

There will be an increase in the percentage of nominal exophoric substitution out of all substitution. Ellipsis can be seen in certain aspects of special terms, as I have shown. An extreme form of substitution and ellipsis is the unfinished sentence. I hypothesise that there will be an increase in sentences ending with a substitute such as 'etc.' or 'and so on,' and sentences which are left incomplete.

e.g. JAN 20: 'So that if you don't get it...'
'Heh heh heh heh.'

As all these contextualisation cues increase, there will be a decrease in post-head dependents (modifiers and peripheral dependents), a reduction in restrictive modification as the bald noun-phrase becomes all that is necessary to identify the referent. An outsider might feel the dialogues inaccessible because of the lack of post-head dependents.

e.g. MAY 12: 'Your CV and your proposal.' (outsider: 'for what?')

My general hypothesis about the contextualisation cues is that after the beginning of the course, there will be an increase in special terms but that as students become more familiar with them, they will use them more loosely and refer to linguistics and course-related referents in more general terms. That is to say, initially there will be a peak of special terms, proper names, demonstrative and comparative reference, combined with a drop in post-head dependents. As the course progresses, special terms and names will level off and there will be an increase in third person personals, exophoric substitution and ellipsis, and superordinates, general words and popular general expressions. This overall trend will be affected by events in the course: I predict minor increases in special terms around exam and portfolio dates and project deadlines.

To complete the study of cues and knowledge areas, I shall take into account two secondary but essential factors: cohesion and function. A consideration of cohesion will reveal that as reference, substitution and ellipsis become more exophoric, lexis more course-related and general, and post-head dependents scarce, the risk of communication breakdowns, or at least requests for clarification, increases. I foresee a greater increase in breakdowns and clarification requests in course-related topics than in non-course-related ones.

The analysis of the function of utterances containing cues is significant because, as Levinson (1978) has shown, claiming common ground and in-group membership, referring to a shared situational context, has a social cohesive effect. The interactional utterances may be a sociable but serious exchange of information to enlighten or an anxious test of the normality of a situation, or they may be a light-hearted relieving of tension with conversational implicature, flouting the maxim of quality, amusing colleagues with joking, irony and banter and interest-holders such as hyperbole and metaphor. I hope to show that the use of contextualisation cues is a generally expected unmarked means of claiming in-group membership.

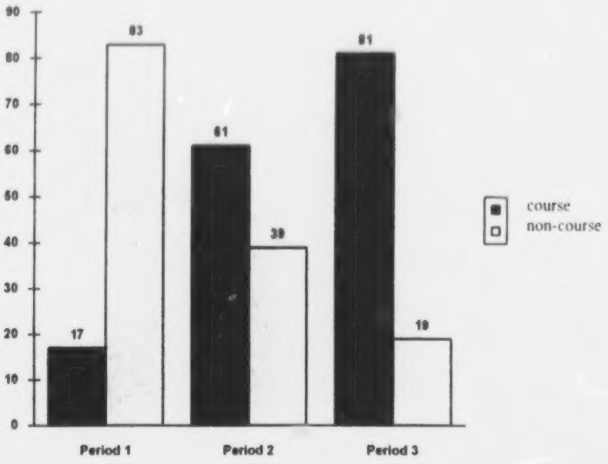
3.2 Method of data collection.

I openly made tape-recordings (total 264 minutes) of MSc student conversations in the common room of the Applied Linguistics department from 4 October 1991 until 12 May 1992. I recorded once a week for the first half of the first, second and third terms. The conversations were spontaneous and unguided, and I kept at a distance at the moment of recording so as not to be included. Six native speaking students who had options in common and tended to sit together in the common room consistently were eventually selected for analysis.

I chose three-minute segments from dialogues in which the greatest number of the six selected students took centre stage together in order to make comparisons easier and more systematic.

Before analysing individual contextualisation cues in these three-minute segments, it was necessary to check that the segments were a representative sample of the c topic: n-c topic ratio of each period. To do this, I calculated the time spent talking on c and n-c topics within each recording, and Figure (2) shows this percentage as an average per period. There is a noticeable increase in c topic time and decrease in n-c topic time. Finally, I calculated the average percentage for each topic type per term in the three-minute segments, and found that the ratio was similar enough to that for all recorded material for any observations about cues to be representative of all the recorded data.

Fig. (2) Average percentage, per period, of time in all recorded data on course and non-course topics.



4. The study of special terms

4.1 Data analysis.

The first problem that arose was that 'special terms' was an inadequate heading, as many terms were not intrinsically 'specialised' but became specialised by their context. It was their pragmatic reference and the previously established schemata that determined which topic type the term belonged to, which K area was being tapped. Thus, for example, I began to feel the need to accept terms such as 'discussion' on one occasion because it was about research methods, and 'this week' on another because it referred to activities in part of the course. However, accepting these terms, which are not intrinsically specialised, seemed to be diluting my argument about an increase in special terms.

I therefore re-defined my 'special terms' categories. Within the macro-category 'special terms' I made four divisions. I now adopted the name 'technical terms' for the category of intrinsically specialised terms independent of context, technical words of linguistics and language teaching such as 'discourse,' 'creoles' and 'lesson plan.' Then I devised a 'c term' category for terms only specialised by context, but intrinsically course-related: 'specific c terms' for ones such as 'core project,' 'portfolio' and 'topic sheet,' 'general c term' for 'syntax' as in 'how's your syntax?', and 'superordinate c term' for the likes of 'book' as in 'have you read the book?' Then I made a second category, 'c-cxt term,' for those terms that are not intrinsically course-related but become course-related by their context, such as 'discussion' and 'this week.' All other terms were obviously 'n-c terms,' or non-course-related, not even by context. Figure (3) shows each knowledge area with the hypothesised principal types of term that are mostly found in it, by definition. This is not to say that the other types cannot occur in each knowledge area, of course.

Fig. (3) General tendency of terms in each knowledge area.

	Special Terms				C-cxt Terms	N-c Terms
	Technical Terms	C Terms				
		Specific	General	Superord.		
K1						X
K1 (C)					X	
K2	X					
K3		X	X	X		
K4		X	X	X		
K5						X

The second problem that arose was that counting terms in spontaneous spoken data is not so straightforward as counting terms in written data. Because spoken data is interactive and unplanned, the same term may be repeated several times.

e.g. MAY 12: DM: Are you talking about a project?
AF: Yeah.
DM: You're talking about a project.
AF: I'm talking about a project.

Speakers repeat interlocutors' words to show solidarity, check comprehension and negotiate meaning. They repeat their own words as they hesitate, think what they are going to say and reformulate their own ideas. I decided to count each occurrence of a noun, even when the same noun was repeated in quick succession as a stutter, because it was difficult to formulate an non-arbitrary rule for counting a repeated word as one, two or three, and I felt that the overall distribution of nouns would not be unbalanced by counting every occurrence. This point is especially important to remember when I measure density.

The third problem was that it was obvious that if c topics themselves become more frequent at the expense of n-c topics, then this should affect the total number of special terms, and that it was not so much the number or frequency of certain terms but in fact their lexical density that might change with time. I therefore also calculated lexical density within each topic type (course-related and non-course-related). Expressions containing 'thing' and proper nouns were not considered, at this stage.

4.2 Results and discussion

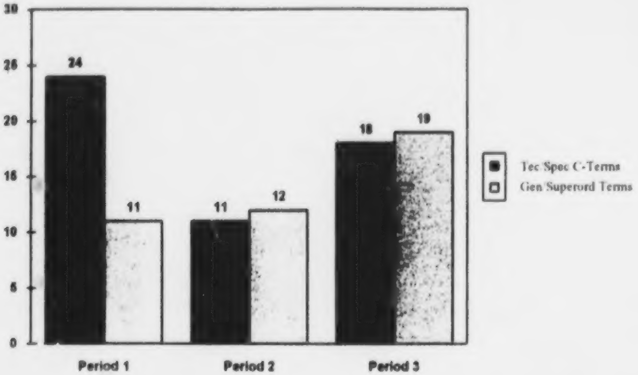
The total number of special terms is shown in Figure (4). The number in the last period was slightly greater than that in the others, probably explained by the increase in c topics time in that period. There were twice the number of total K3 and K4 c terms than K2 technical terms.

Fig. (4) Total number of special terms in K2, K3, and K4.

Period	K2	K3	K4	TOTAL
Oct/Nov	3	2	30	35
Jan/Feb	4	17	8	29
Apr/May	10	24	5	39
TOTAL	17	43	43	103

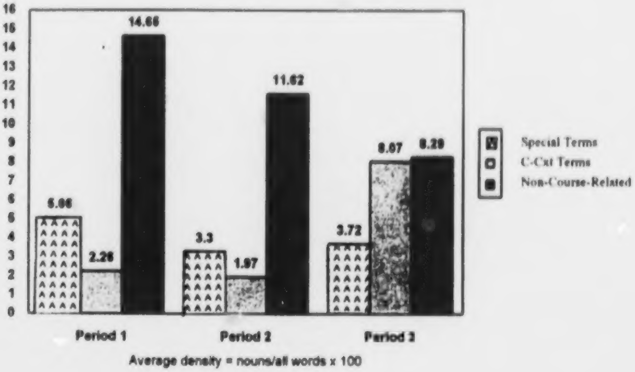
In order to discover how each special term increases or decreases over the three periods recorded, I examined each of the four categories individually over time. Figure (5) lays out these developments over time graphically and shows that there are more technical and specific c terms at the beginning of the course and more general and superordinate c terms at the end. This suggests that my hypothesis is confirmed but it may not be a very reliable calculation given the size of the sample. The same calculations would need to be made with all the recorded data for this to have more significance. Whereas all special terms in K2 are simply technical ones, special terms in K3 and K4 can be any of the c term types. In K3, they are as often specific c terms such as 'reading week,' 'PhD' as superordinate c terms such as 'class,' 'option,' 'project.' About 20% are general c terms and in all cases they refer to courses: 'language and linguistics,' 'psycho-linguistics.' K4 contains fewer superordinates (e.g. 'group,' 'questions,' 'books') than K3 does but the same number of general c terms, and these are in-group names and abbreviations such as 'Psycho', 'Teap,' etc.

Fig. (5) Totals of special term types for each period.



I then calculated the density of special terms and c-cxt terms in c topics and the density of n-c terms in n-c topics. By density, I mean the percentage of nouns of a particular type out of all the words in one topic type. Figure (6) shows that while special term density remains constant, c-cxt terms increase in density. The density of n-c terms in n-c topics decreases, suggesting that while knowledge assumed in the course context is soon established, it takes longer for n-c topic knowledge to be taken-for-granted.

Fig. (6) Average density of various term types per period within each topic type.



Finally, all these calculations must be seen against the background of the discoveries made about the increase in time spent on course topics mentioned in 2.2.2. The inaccessibility of the dialogues may well be because of the greater proportion of time spent on course topics with a consistently low lexical density of special terms, which are themselves increasingly general.

4.3 Further research

It is clear that this analysis should be applied to all the recorded data, as the three minute segments do not contain enough examples of each type of term for the results to be reliable. Two more quantitative tests need to be done with special terms. The first is to observe how they behave intertextually, how they thread through the various dialogues over the whole year. The second is to examine the rest of the noun phrase: whether it is definite or indefinite, how it is modified, etc., to discover how the terms are actually used.

What now need to be examined more closely are other factors affecting the use of special terms. One obvious factor is that of deadlines within the structure of the course programme. There appears to be an increase in technical and specific c terms around exam-time and the dates for handing in projects.

In addition, special terms need to be analysed from a qualitative and functional point of view. Questions to be examined are: are special terms used as a marker of group identity, as a demonstration of in-groupness, of solidarity with interlocutors? Or are they used to test whether the progress of others is the same as that of the speakers? Technical terms are used relatively little in the common room: how do interlocutors react to those who use them freely? The function of the use of in-group language cannot be readily determined in a satisfying way. As an initial step, I shall attempt to carry out a bottom-up survey, classifying each discourse unit in terms of speech acts and moves. In the pilot study, triangulation interviews with recordees elicited global macro-functional comments of an unquantifiable nature.

Finally, the question of how to measure inaccessibility to outsiders, impenetrability of conversations, needs examination. I have devised questionnaires based on four of the recorded dialogues. The questionnaires contain questions of general comprehension (knowledge) and specific understanding of isolated words (contextualisation cues). Those who fill in the questionnaires will possess one or more of the five knowledge areas: that expected of a non-language teacher, that of a member of staff of the department of Applied Linguistics, that of the MSc students recorded etc. I hypothesise that those closest to the 1991-92 MSc group will best be able to unravel the reference of contextualisation cues and that this makes the conversations penetrable.

5. Conclusions

This paper has suggested a model for analysing changes in casual conversations of students as they form a discourse community. Beginning with the observation that the literature on the difference between the language of strangers/acquaintances and that of friends does not describe the process of passing from one stage to the other, this paper has offered hypotheses of a lexical and grammatical pragmatic nature.

16

The second part of the paper has consisted of a brief exploration of the first of the lexical hypotheses: that special terms increase, or rather that technical and specific course-related terms increase at the beginning while general and superordinate course-related terms do not emerge and increase until the second period of recording, that is, half way through the course, when assumed shared knowledge about the context of the course has grown and reference to it needs to be less explicit.

Once all the hypotheses have been tested and seen together, cohesion and function examined, and a statement about group formation made, I hope to explore the pedagogical implications. There is still a need for courses to train learners to guess what exophoric reference, substitution, ellipsis and general words might refer to, taking into account that even native speakers have difficulties with such reference. This approach could be particularly useful for EAP students, as it might help them to understand and participate in conversations between native speakers of English in their departments. Using the information that this study should produce about the relationship between grammatical and lexical cues and about which are most frequent in what type of knowledge area and for what function, materials could be devised to train learners to use bottom-up procedures to use the special lexis, terms and names to build up their own picture of a possible presupposition pool of contextual knowledge, and then from there to use top-down procedures to guess what part of the schemata the general words, reference, substitution and ellipsis might refer to.

Learners could also be trained to ascertain whether a dialogue is between strangers or between in-group members of a discourse community, by looking for the cues. They could also be trained to appreciate whether in-group members have recently entered the community or whether they have been in for longer, by looking for general and elliptical expressions.

References

Altman I. and D. Taylor. 1973. Social Penetration: the Development of Interpersonal Relationships. New York: Holt Rinehart and Winston.

Argyle M. and P. Trower. 1979. Person to Person - Ways of Communicating. Holland: Multimedia Publications.

Berger C. and J. Bradac. 1982. Language and Social Knowledge. London: Edward Arnold Ltd.

Bernstein B. 1971. Class, Codes and Control (Vol I). London: Routledge and Kegan Paul.

Brown P. and S. Levinson. 1978. Politeness. Cambridge: CUP.

Cook G. 1989. Discourse. Oxford: OUP.

Coulthard M. 1985. An Introduction to Discourse Analysis. London: Longman.

Giles H. and R. St. Clair. 1979. Language and Social Psychology. Oxford: Basil Blackwell.

Gumperz J. 1982. Discourse Strategies. Cambridge: CUP.

Halliday M. and R. Hasan. 1976. Cohesion in English. London: Longman.

Hinde R. 1987. Individuals, Relationships and Culture. Cambridge: CUP.

Joos M. 1962. 'The five clocks.' in Folklore and Linguistics. Bloomington, Indiana: University Research Center in Anthropology.

Kelly H., E. Bersheid, A. Christensen, et al. 1983. Close Relationships. New York: W.H. Freeman and Company.

Kreckel M. 1981. Communicative Arts and Shared Knowledge in Natural Discourse. London: Academic Press.

Levinson S. 1983. Pragmatics. Cambridge: CUP.

Levy D. 1979. 'Communicative goals and strategies: between discourse and syntax'. in Given, T. (ed.) Syntax and Semantics 12:183-210.

McCarthy M. 1991. Discourse Analysis for Language Teachers. Cambridge: CUP.

Swales J. 1990. Genre Analysis. Cambridge: CUP.

Tannen D. 1984. Conversational Style: Analyzing Talk Among Friends. New Jersey: Ablex Publishing Corporation.

Tannen D. 1989. Talking Voices. Cambridge: CUP. 22

SPECULATION AND EMPIRICISM IN APPLIED LINGUISTICS[1]

Alan Davies (DAL)

Abstract

Speculation ought not to be a pejorative term, and ought not to be in conflict with empiricism. Davies contrasts two traditions, one originally seeking applications for theory, the other looking for solutions to problems in FLT, and finds both valuable. Five applied linguistics topics - curriculum, discourse analysis, systemic linguistics, testing, second language acquisition - are briefly discussed within this framework. Davies concludes, in broad agreement with Widdowson, that the value of empirical research depends upon the quality of conceptual analysis, and advocates scepticism and humility.

> 'Thy bones are marrowless, thy blood is cold,
> Thou hast no speculation in those eyes
> Which thou dost glare with'
> (Macbeth to Banquo's Ghost, Macbeth 3.4:93)

0. It is a common criticism of applied linguistics - a criticism made by its practitioners as much as anyone - that there is no objectivity about it, that its views and hypotheses and conclusions are determined by fashion rather than by rigorous scientific procedure, that in fact there are no hard data because there is no way of establishing whether something is a result or a finding. This is a two-fold criticism. It is a theoretical criticism, denying that applied linguistics has any organised body of theory, and it is an experimental criticism, arguing that even if there is any body of theory there is no link between that and arguments as to how to proceed, i.e. how to teach and learn languages. As a result, in language teaching as in education generally, what determines change is the roundabout of fashion which seems recently to be moving back towards a modified grammar-translation method after a number of years in which such an approach to language teaching was anathema to many people. It may be that we shall always have to take account of changing fashion simply because we have no way of finally establishing 'the best way' to learn or teach a language. Since there is no easy way of evaluating the internal logic of a theoretical model of language, the question of what constitutes the best language-learning theory may not be a matter for experimental research at all, but a matter for philosophical argument about what kinds of aims we are interested in at any one time. Doubtless these will be influenced by...within-theory experimentation...our only hope of escaping from the tyranny of fashion is through submitting our guess-work to the rigour of hypothesis and experimentation (Davies 1977:1)

---

[1] A version of this paper was read at the 16th annual meeting of the Applied Linguistic Association of Australia in Townsville, September/October 1991.

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS, number 4 (1993) ISSN 0959-2253

I wrote that comment in the mid-70s. It was published as the opening to the Introduction to Volume 4 of the Edinburgh Course in Applied Linguistics (Allen and Davies 1977). I quote my own words, partly of course out of astonishment that that was actually me all those years ago, but more importantly because I want to suggest that our unease about the role of fashion, about the apparent tensions between speculating and empirical requirements and imperatives, that these are just not new, that they have always been around.

They are more important in my view than that weasel-like dismissal of applied linguistics, that it lacks coherence. Yes indeed, I admit it, it does lack coherence. But I just do not see that as negative; indeed it does not make applied linguistics any different from any other academic discipline that I know of, linguistics, English, education, even medicine and law. They are all loose federations, often warring ones more on the model of Yugoslavia than of Australia or the European Community, but in no case is there a single monolithic, unitary view, nowhere is there complete agreement of what the discipline is about. No, academic disciplines, certainly academic departments, are political groupings, which of course means that over time it is proper for them to regroup. Of course there are some interests that are closer and some that are further apart. In that context applied linguistics is actually in a strong position, if only because it is centrally about language, about intervention in language problems (such as in teaching) and about language treatment (such as language planning). In terms of social and human focus applied linguistics is in as strong and as coherent a position as is, say, medicine. As the title of this paper indicates, speculation and empiricism do concern me, not that they are in some sort of conflict or tension, no, because again that appears to be normal for academic disciplines, but because we are unhappy about their coevality. We should not be. They are both there, they are both necessary and we should welcome their presence as our discipline matures.

I think that was what I was trying to say back in 1977, that speculation and empiricism both had their place, and as such were capable of generating both philosophical argument and the rigour of hypothesis and experimentation.

1. Speculation seems to have fallen into bad company. From the sense of 'contemplation, consideration or profound study of some subject' and 'conclusion reached by abstract or hypothetical reasoning' it has come to be used in somewhat disparaging ways, often preceded by 'mere', 'bare' or 'pure', implying conjecture or surmise. This of course quite apart from its more operatic senses of 'action or practice of buying and selling goods, lands, stocks and shares etc. in order to profit by the rise or fall in the market value as distinct from regular trading or investment; engagement in any business enterprise or transaction of a venturesome or risky nature, but offering the chance of great or unusual gain'. Alas! try as they may no applied linguistic speculator has, as far as I am aware, yet reached great or unusual gain though there are rumours that Stephen Krashen has put in a joint bid for The Age with Packer!

There is also the sense in speculation (though it is not made explicitly) of some deductive process. That of course matches the inductive label attached to empiricism, which is defined as 'the use of empirical methods in any art or science', empirical itself receiving rather shorter shrift as having a concern for observation and experience more than for theory ('derived from or guided by experience or experiment; depending upon experience or observation alone, without using science or theory'-Macquarie Dictionary).

It turns out that speculation and empiricism should not in fact be in conflict. What contradicts empiricism is rationalism. While 'empiricism' attracts the comment:

> reason cannot of its own provide us with knowledge of reality without reference to sense experience and the use of our sense organs. (Angeles 1981).

'rationalism' has this one:

> reality is knowable...independently of observation, experience and the use of empirical methods; reason is the principal organ of knowledge and science is basically a rationally conceived deductive system only indirectly concerned with sense experience. (ibid.)

It would be convenient to agree that speculation combines the two senses of (random) conjecture and of reasoning attaching to some explanatory theory, while empiricism means the use of experimental methods to validate a theory. However, what seems to have happened is that empirical has appropriated to itself the package of the scientific methods, theory *plus* controlled enquiry, while speculation has increasingly been marginalised to the armchair, the haphazard and the guess.

What, after all, of this definition:

> speculative philosophy: in the pejorative sense: philosophy which constructs idle thoughts about idle subjects? (ibid.).

Happily, speculation is not only this snapper-up of unconsidered trifles. In the same definition of speculative philosophy, we read:

> in the non-pejorative sense: philosophy which constructs a synthesis of knowledge from many fields (the sciences, the arts, religion, ethics, social sciences) and theorizes (reflects) about such things as its significance to humankind, and about what it indicates about reality as a whole. (ibid.)

It is of course the case that, for many, speculation in this latter sense remains a noble activity; and I shall argue that in applied linguistics we need both speculation and empiricism; indeed, one of the characteristics of applied linguistics is that even hunches or guesses always come from somewhere. Just like other academic areas. And when mere speculation in applied linguistics is once again being held up to scorn because it is not experimental it will be well to remember that Macbeth's criterion for Banquo's being a ghost, for his not being alive, was precisely that he lacked speculation:

> hast no speculation in those eyes
> Which thou dost glare with'

Ernest Gellner quotes Keynes:

> "The ideas of economists and political philosophers...are more powerful that is commonly understood. Indeed the world is ruled by little else. Practical men who believe themselves to be quite exempt from any intellectual influences are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back". This is true far beyond the sphere of economic thought. Those who spurn philosophical history are slaves of defunct thinkers and unexamined theories. (Gellner 1991:11,12).

2. If the ontogenesis of an academic or scientific discipline has any phylogenetic status, then we might posit that as it matures it becomes increasingly empirical but does not cease to be speculative. Just as human societies show a movement from hunting-gathering (where change is wholly evolutionary) through the agrarian (where change is by choice), so maturing disciplines move towards a deliberate marriage between the speculative and the experimental so as to make what is investigable what is also worth investigating. It seems to be a characteristic of a poor experimenter as of non-serious discipline that their research questions are unresearchable. 'Industria' (Gellner's name for the stage of industrial society)

> is not based on any one discovery, but rather on the generic or second-hand discovery that successful systematic investigation of Nature, and the application of the findings for the purpose of increased output, are feasible, and, once initiated, not too difficult. (ibid: 17-18)

We might perhaps make that a criterion of a mature discipline and by that token ask ourselves whether applied linguistics contains that successful systematic investigation of Nature, increased output (interpreted as we will), and whether the systematic investigation is not itself too difficult.

3. It will be helpful to consider two opposing applied linguistic traditions, both of which are still very much influencing what we teach, what we research and how we see ourselves as applied linguists. One model starts with theory (typically linguistic theory), the other with practice. The first (Linguistics-Applied) has had much influence in North America and in Continental Europe (and I think also in Australia); the second (Applied-Linguistics) is more commonly found in Britain and some parts of the Commonwealth. The American Linguistics-Applied tradition starts with linguistic theory and looks for ways to apply it most usefully on practical problems such as language teaching; the British Applied-Linguistics tradition starts with the practical problems and then seeks theoretical (and/or practical) ways to understand and resolve those problems. The North American tradition of Linguistics-Applied grew out of the search by linguists (e.g. Bloomfield, Fried) for applications for their theoretical and descriptive interests. These applications they found in language teaching, especially during the Second World War. The foundation of the English Language Institute at Ann Arbor, Michigan, was one of the key initiatives in American applied linguistics, representing a substantial intellectual investment in language teaching by linguists, either faculty members or graduate students, whose chief interest was in the main in linguistics not in language teaching. American applied linguistics can therefore be characterised as Linguistics-Applied, an essentially

top-down approach. This tradition also holds in Britain in the work for example of J.R. Firth (also very much involved during World War Two in intensive language teaching courses), and of his student, Michael Halliday; hence of course my comment above about the situation in Australia, at least in the beginnings of applied linguistics here.

It is now however the mainstream British and Commonwealth tradition, which comes from quite a different source, that of teaching English as a foreign/second language in the former colonies, in Latin America, Japan and Continental Europe, above all outside the UK (and here may be another link between the Australian and the North American experience). The work that the British Council took on under Arthur King and developed widely around the world was in this tradition of professionalising language teaching. It was very much a bottom-up approach to the field and it led inevitably to a search for input of a theoretical kind. Hence the establishment in 1957 of the School of Applied Linguistics in the University of Edinburgh precisely to provide that theoretical backing and support.

For over 20 years from 1964 Pit Corder directed that effort. It is significant for our argument that in his own writing and scholarship Corder eventually found that tradition incoherent in its attempt to marry bits of theory to practical issues. What Corder's case indicates is that reliance on one or other of the two traditions alone (Applied-Linguistics and Linguistics-Applied) is inadequate: in his case a career which was so much in the mainstream of British applied linguistics and so successful in directing it needed to break with that tradition in order to make his major contribution, in the concept of interlanguage.

Corder's model of second language acquisition, interlanguage, based on speculation, has a stronger claim than most to be called a theory. For seriously empirical colleagues in North America and Europe, it never mattered that Corder's work was not empirical. For them he was the theory maker, and if there does now exist a theory of interlanguage and of Second Language Acquisition (SLA) then it is because of Corder's thinking and writing about these issues. He never disdained the label 'speculation', acknowledging that speculation necessarily antedates the empirical work that leads to the development of theory.

4. Brumfit (1985) suggests three types of applied linguistics research: policy oriented, truth seeking and action. Examples of the first might be curriculum study, of the second SLA and of the third test construction. I am however not easy about Brumfit's three-way split and would prefer a binary division: truth seeking and action/policy oriented. I shall call truth-seeking explanation and policy oriented/action practice. I hope this scheme will help make sense of both the teaching and the research aspects of applied linguistics courses. In addition to explanation and practice outcomes there is a third type of research dynamic, evaluation. For the present, however, it seems to me more helpful to regard evaluation as one aspect of the practice type of research.

5. Can the same be said of teaching applied linguistics? I want to propose that teaching and research in applied linguistics have similar purposes. Both are concerned with explanation, one with its expansion, the other with its dissemination; both are concerned with evaluation, one through particular types of research (we suggested policy oriented), the other through assessment of teaching and learning; both are also

concerned with practical outcomes, where research may be seeking new insights and solutions and teaching is training teachers etc. to implement those insights.

I propose therefore this matrix:

	Research	Teaching
Explanation	/	/
Practice: Problems	/	
Practice: Skills		/

Evaluation, we should remember, is incorporated in both Problems and Skills. That separation into Problems and Skills is in any case opportunistic as a way of stating something of the obvious about the difference between teaching and research.

Now to my five topics in applied linguistics, in each case offering a priority order as between Explanation and Problems/Skills, my hypothesis being that what determines that priority is not primarily the research-teaching distinction but something else, perhaps the contemporary urgency of the topic. We may ask ourselves of course whether, in terms of our earlier discussion about the maturing of disciplines, we would expect the basic division to be between teaching and research such that research is - by definition - primarily explanation oriented and teaching basically practice/skills oriented. We will return to this question.

The five topics I survey are all researched and taught within applied linguistics programmes. Three (curriculum, discourse analysis and systemic linguistics) I will approach from the point of view of their involvement as components of applied linguistics course work while the other two (language testing and second language acquisition) I will discuss from the point of view of their research capabilities. By this selection I make no statement about the priority or otherwise of these five topics within applied linguistics, nor do I imply any value judgements among the five selected about which ones are more important in research or in teaching. My choice to discuss some in their research context and some in their teaching context is arbitrary, happenstance.

6. Teaching

6.1 Curriculum

White (1988) offers three models for change in an existing curriculum:

1. research/development and diffusion/dissemination

2. problem-solving

3. social interaction.

He then examines three types of innovation strategy: power-coercive, empirical-rational and normative-re-educative. His conclusion is as follows:

> On the whole...innovations which are identified by the users themselves (rather than specified by an outside change agent) will be more effectively and durably installed than those which are imported from outside, since it is the teachers and students themselves who will have 'ownership' of and commitment to the innovation concerned if it has a grass roots or bottom-up rather than a top-down origin. For this reason, a problem-solving model and a normative-re-educative approach to innovation will probably be the most successful combination in language teaching as elsewhere (White 1988:133).

This problem-solving philosophy is sometimes associated with Stenhouse (1975) and his ideas of action research. As such the lines between skills and explanations are elided. While I agree with White about the effectiveness of grass roots change I am not as sanguine as he appears to be about the necessary attitude change taking place from within. 'Normative change will involve alteration in attitudes, values, skills and significant relationships' (White 1988:129). True, he does also point out that 'direct interventions by change agents' are necessary. But what is really being asked for here is a sophisticated language teaching culture (to go right back to the beginning of this paper, a choice or industrial culture) which is difficult to create *ab initio*. It is as though curriculum is the last topic to expect change in rather than, as is so often the case, the first to be enlisted.

Widdowson (1990) is helpfully outspoken on the role of empirical research in determining language teaching outcomes (one of the problems with Widdowson's position is that he switches backwards and forwards in his discussion between language teaching and applied linguistics). For Widdowson empirical research has nothing to offer language teaching in terms of solutions. His view is that what is needed is appropriate conceptualisation. There is, he suggests, in discussing Krashen, a need for clear thinking.

Widdowson overreaches himself, first, because of the Gellner argument about latent scholarly influences and, second, because his denial of the possibility of pedagogical problems being solved simply because there is vaguely relevant empirical research is effectually an Aunt Sally, an *ignaratio elencti*. When Widdowson is concerned with syllabus (curriculum, as White points out they are used interchangeably) his conclusion is that it is 'unlikely that any research at present or in the future will provide us with anything very definite to resolve these difficulties' (1990: 154). What matter for Widdowson are first that the principles on which the syllabus has been designed are explicit and second that the teachers should be methodologically aware. But this is surely sleight of hand. No matter what we call it, curriculum or syllabus, or syllabus and/or methodology, there is always a delivery issue for language teaching, and the problem surely is how to provide for that delivery. My own view is that curriculum/syllabus/methodology is always problem-oriented and that there is also a necessary secondary research (explanatory) aspect.

6.2 Discourse analysis

Much, perhaps most analysis of linguistic systems including discourse makes use of data. No doubt, as with novels, however invented the examples of spoken and written texts and interactions might be, they still to some extent relate to reality, but of course it is a question of how close. True, the idealised conversations we find in invented texts such as novels are based upon the writer's knowledge of the language but, as we also know, that knowledge is diverse. In other words, what the writer invents or imagines may tell us only about the writer's invention and imagination, not about what s/he actually says and how s/he actually behaves in daily life. It is of course an extreme form of the observer's paradox.

For some linguists this is no problem. Gazdar maintains 'I shall assume...that invented strings and certain intuitive judgements about them constitute legitimate data for linguistic research'. (Gazdar 1979:11 quoted in Brown and Yule 1983:20).

Brown and Yule themselves take a different view, and in my opinion the correct one. Their material, they claim 'is typically based on the linguistic output of someone other than the analyst' (1983:20). They summarise their approach as follows:

> the discourse analyst treats his data as the record (text) of a dynamic process in which language was used as an instrument of communication in a context by a speaker/writer to express meanings and achieve intentions (discourse). Working from this data, the analyst seeks to describe regularities in the linguistic realisations used by people to communicate those meanings and intentions. (1983:26)

Guy Cook, describing the Birmingham discourse analysis 'school' in his recent book on Discourse (Cook 1989), tells us:

> Sinclair and Coulthard recorded a number of British primary school lessons. On the basis of these data they proposed a rank structure for these lessons as follows...They then drew upon rules based on the data. (Cook 1989:46-7).

Whatever we may think of the Birmingham school and even though we know that the primary school lessons they recorded contained at most only 8 children in each lesson to make the recording easier, we have to accept that foremost in discourse analysis research is explanation, and that in teaching discourse we are in our applied linguistics classes more concerned with disseminating what is known about discourse than about how to do it. Of course, as with teaching about grammar (or indeed statistics) it is surely the case that for some learners operating new skills and expanding their knowledge go hand in hand, that a totally conceptual approach is ineffective, that it needs to be accompanied by a skills (how to do discourse analysis) workshop. At the same time, what is primary is surely the dissemination of the knowledge, and even if we are teaching it hands-on that is because we are primarily concerned with the knowledge not the skills.

6.3 Systemic linguistics

As we have already noted, systemic linguistics (or systemic functional linguistics, or to use an earlier term scale and category grammar, or in one of its Australian applications genre theory) has been very influential in Australian applied linguistics and in particular in Australian educational linguistics. For those of us who are not systemicists there is a real problem of relativity in trying to come to terms with the totality of approach that seems to be required for systemicists. Let me give a personal example. Some years ago after a reorganisation of departments in the university I was working in I asked one of my recently acquired colleagues what she was interested in teaching. Her reply was both generous and at the same time obscurantist. 'I'll teach anything you like, Alan, but you must remember I'm a systemicist!' Game set and match to her! Like a religion, I thought to myself, impossible to argue against.

Be that as it may, its influence has as I have suggested been profound in Australia. And the influence, even though it is so largely in schools, is not so much in the teaching of skills. Like discourse analysis its mission is really about knowledge. As Martin quite charmingly remarks:

> You could accuse me, like everyone else, of being after power. I want people to see that the way a linguist looks at language makes explicit what we implicitly know and explains why we do often act as we do. (Martin 1985:62)

Well that's a fair cop! That really does seem to put systemics into the explanation bag rather than the skills bag, doesn't it?

I have mentioned relativity and its echoes of Whorf. I suppose that Whorfianism inevitably assumes a God's truth perspective. (And how interesting it is that both systemics and generative linguistics share that pursuit which for the rest of us is sadly the hunting of yet another snark.)

> It will be argued that genre, or to use the more technical linguistic term, register, exists. (Martin 1980:1)

Note that this register-genre distinction, which I have always found quite recondite, is here apparently non-existent. But the existence of genre (or register) is a given.

> There can be no doubt that genres exist; but exactly what they are is the subject of another generation's or two's research. (Martin and Rothery 1981:50).

Is it disingenuous of me to find something odd about that sentence? Genres exist but we don't know what they are! Something rather alien perhaps. Surely if you are so certain they exist then it must be possible to determine similarities of shape, behaviour and so on.

It seems that progress in Sydney (and Geelong?) was faster than anticipated and well before 'another generation or two' the truth had begun to emerge, so that Rothery could write:

> Teachers have always been aware of different varieties of writing. Narrative, Report and Exposition are commonly asked for in school. But what they have not been aware of is that the organisation or stages of these texts can be identified in distinctive ways and this is what constitutes a text's genre. (Rothery 1986:117).

We may find the actual analysis of genre (narrative has three stages: orientation, complication and resolution) somewhat flimsy, but for our purposes here that is beside the point. Clearly the systemic agenda is to impart knowledge about genre and as an adjunct to help teachers develop the skill of genre construction in their pupils.

The basic argument of genre apologists seems to me irrefutable, it is that 'genres are learned' (Rothery 1986:123). There seem to me to be two problems which constantly get in the way of this very sensible message. The first is that there appears to be a vendetta against one Donald Graves, who is stalking the land preaching his heresy of process writing, and who is reported as very bad news indeed; the other is that the pursuit and now the actualising of genres seems to me not only wrong but unnecessary. As with variety, as with register, so with genre; sure language has variety, it has register, it has genre, but that is not to say (indeed it is snark-like to pretend it can be demonstrated) that there are varieties, registers, genres which are describable, separate and discrete. It is of course the old language problem (rather languages problem) under another name.

By 1987 the work had gone on apace and there is now greater clarity about the distinction between genre and register (a distinction which you will remember was non-existent in 1980.):

> genre theory differs from register theory in the amount of emphasis it places on social purpose as a determining variable in language use' (Martin, Christie and Rothery 1987:119-20).

Taking three examples of topics within applied linguistics teaching programmes, I have argued that in each of the three cases Explanation and Practice play a part; further that in individual cases the emphasis is likely to be on one more than on the other. In Curriculum it is more on Practice, while in discourse analysis and systemics it is more on Explanation. I turn now briefly to two topics from research directions in applied linguistics, Testing and Second Language Acquisition (SLA). They make an interesting comparison pair in that they appear to have started from quite different origins and have moved in the last 15 years in opposite directions to one another. As we shall see that is very much a simplification.

7. Research

7.1 Language testing

Language testing is the prime example in research of being (or of having been) at the developmental end. It exists as it were to create new tests, trialled, validated and so on, but nevertheless not originating new ideas about language definitions or learning. That has now partly changed since language testing has in the last years come to be at the cutting edge of our investigations into proficiency (the unitary nature of language), the meaning of the native speaker, the definition of communicative competence, as well as questions about variety (the status of languages for specific purposes) (Davies 1990). What is more, language testing has developed new methodologies or at least made use of alternative ones for its own investigations, always a sign of a maturing discipline. And yet I would want to say that language testing in its research mode is still primarily a practice (problems) research discipline.

7.2 Second language acquisition

If we can for the moment forget about the error analysis origin of SLA, an origin like some forms of poverty and obscurity in birth which is also conveniently forgotten, then SLA (as we saw in our discussion about Corder's interlanguage), like that other three letter word sex, began in the 1960s, as a deliberate attempt to raise the theoretical stakes in applied linguistics on the analogy of child language acquisition. As we also saw, Corder's initiative was quickly taken up empirically and I would say has been in the last 15 years over-studied empirically. As Widdowson would no doubt say, we have left ourselves too little conceptual analysis, too little explanation, too many trees, too little of the wood. That seems to me now at last to be changing. Larsen-Freeman and Long (1991), hard-nosed empiricists both, comment in their recent survey on how many studies there have been ('a four-fold growth' 1991: 5) but state that there is indeed now a need for more 'research studies which concentrate on improving our understanding of the effect of choosing from among particular instructional design features' (1991:332). That seems to mean that they think SLA research should take more interest in facilitating and expediting the SLA process (1991:6). Nevertheless for my money (as well as Larsen-Freeman and Long's) at the moment it is clear that for most people who regard themselves as SLA researchers (as opposed to researchers into second language learning) it is explanation that has top priority.

8. Matrix

So we can now fill in the matrix we offered in 5.

		Research		Teaching	
	LT	SLA	DA	System.	Curric.
Explanation	L	H	H	H	L
Practice: Problems	H	L			
Practice: Skills			L	L	H

To return to Speculation and Empiricism: just as teaching and research in applied linguistics both contain aspects of explanation and of practice, so too do they both admit of speculation and empiricism which, after all, turn out at best to be ways of, methods of doing scholarship. Over time (as we saw with language testing and SLA) a change in priority may occur, and it may be that there is a natural life cycle of a topic as there is of a discipline. Or it may also be that we need to be more interventionist, more deliberate, such that when a discipline seems to be moving away from an applied interest and becoming self regarding, setting up its own research agendas (SLA until recently, language testing now?), becoming separate from applied linguistics, perhaps than we need to take action. I am however reluctant to suggest what action, since in such cases it may be that what is happening is in itself healthy (and may change again with time, as perhaps in the Error Analysis to SLA and now to second language learning?) If there is still need for the discarded topic, then it may be best to start up a new topic. That after all is why applied linguistics got going in the first place, because linguistics seemed to become less and less interested in language learning and language teaching.

In conclusion I find myself close to the Widdowson view, the primacy of clear thinking and of theory.

> The value of empirical research ultimately depends on the quality of conceptual analysis that defines the objects of enquiry. (Widdowson 1990:25).

Unlike him I am not a complete nominalist since I believe there is such a thing as data, not too much of it and always purposive and within a theoretical framework. But we can afford to relax: applied linguistics, like any other derivative of philosophy, needs both explanations and skills to make its activity worthwhile. Sometimes one will be more important in one area than another. No matter. The five topics I have mentioned seem to me to be engaged in lively debates about the proper balance between the two. Scepticism and humility, those are the two chief scholarly virtues we all need more of.

> An extensive knowledge is needful to thinking people - it takes away the heat and fever; and helps, by widening speculation, to ease the Burden of the Mystery. (Keats, letter to JH Reynolds 3/5/1818)

References

Brown G. and G. Yule. 1983. Discourse Analysis. Cambridge: CUP.

Cook G. 1989. Discourse. Oxford: OUP

Davies A. 1991. 'British Applied Linguistics: the Contribution of S. Pit Corder'.

Davies A. 1990. The Principles of Language Testing. Oxford: Blackwell.

Gellner E. 1991. Plough Sword and Book. London: Paladin.

Larsen-Freeman D. and M. Long. 1991. An Introduction to Second Language Acquisition Research. London: Longman.

Martin J.R., F. Christie and J. Rothery. 1981 'The ontogenesis of written genre.' Working Papers in Linguistics 2. US/DL.

Martin J.R., F. Christie and J. Rothery. 1987. 'Social processes in education: a reply to Sawyer and Watson (and others).' Working Papers in Linguistics 5. US/DL

Martin J.R.F. 1980. Writing project paper 2: exposition: literary criticism. Working Papers in Linguistics 1. US/DL.

Martin J.R. 1985. Factual Writing: Exploring and Challenging Social Reality. Deakin University Press.

Rothery J. 1986. Let's teach children to write. Working Papers in Linguistics 4. US/DL.

White R.V. 1988. The ELT Curriculum. Oxford: Blackwell.

Widdowson H.G. 1990. Aspects of language teaching. Oxford: OUP.

SEA SPEAK

(In memory of Peter Strevens.)

Alan Davies (DAL)

('academies have been instituted, to guard the avenues of their languages, to retain fugitives, and repulse intruders; but their vigilance and activity have hitherto been vain; sounds are too volatile and subtile for legal restraints; to enchain syllables, and to lash the wind, are equally the undertakings of pride, unwilling to measure its desires by its strength. It is remarkable that, in reviewing my collection, I found the word SEA unexemplified'. Samuel Johnson, Preface to a Dictionary of the English Language, 1755)

In Welsh there is no word for blue
Or green, just one for both must do.
So when I hear the boasts for green
I'm glad it's really blue they mean.

Farm grass for food, drink rain from tree,
Green earth, blue sky breathe life from sea.
Life's basic stuff, land's waiting crowd,
Fond names we rarely speak out loud.

Court orders may preserve one tree,
Living alone protects the sea.
To lash the wind gives pride to names,
Reducing life to language games.

IMPLEMENTING INNOVATION IN LANGUAGE EDUCATION

Gibson Ferguson (IALS)

Abstract

This paper offers a commentary on the problems of implementing innovation with particular reference to ELT. Various factors that influence adoption and implementation are considered: properties of the innovation, the transmission process, and the management of change. The overall aim is to contribute to a sounder conceptualisation of the change process which will assist those involved in the management of change.

Change and its implementation is a topic that has attracted increased attention from the ELT profession (see White 1988, Kennedy 1988, Woods 1988, British Council 1989, 1990, 1991) . This is both unsurprising and welcome.

Unsurprising because ELT professionals are centrally involved in the management of change in various capacities: as teacher educators trying to effect change at an individual or classroom level, as curriculum developers or testers attempting to renew curricula, as managers responsible for innovation in the context of educational aid projects. Given these concerns, it was perhaps inevitable that systematic theoretical and practical enquiry would ensue.

Welcome because it is a corrective to a tendency in the profession to focus overmuch on the content of change at the expense of the process of accomplishing it.

This paper does not, and cannot, review the large literature on change. The purpose rather is to distil from the literature a number of guidelines supported by commentary. The aim is to encourage a sounder conceptualisation of the implementation of change.

1. Innovation: matters of terminology and definition

'Innovation' denotes both a process and a product. By the latter we mean an idea, artefact or practice which is new.

The literature divides the process of innovation into three phases: initiation, implementation, and institutionalization. Initiation is the phase when a problem is identified and a decision to change taken. Resources are then mobilized. In the implementation phase plans for change are formulated and the innovation is put into use. Institutionalization means the incorporation of the new practices into the routines of the institution. The innovation is consolidated.

A similar distinction is sometimes made between adoption and implementation. Adoption is the decision to introduce a particular innovation and implementation the use of that implementation. Most recent research has focused on implementation (Fullan 1989).

This paper is primarily concerned with implementation, though the boundaries between the three phases are not always clearcut.

Innovation as product can be described in terms of two dimensions: depth and scale. First, depth.

Educational innovations may involve any or all of the following levels of change.

a. Structural change: e.g. changes in policy, in timetabling, in grouping of students etc. These largely pertain to the administrative arrangements for instruction.

b. Technological change: e.g. the introduction of computers, video, language laboratories etc. into the instructional process.

c. Materials change: e.g. new books, syllabuses or examinations.

d. Behavioural change: e.g. changes in what teachers do in the classroom, in their teaching style and behaviours.

e. Change in belief, attitude, understanding: e.g. change in teacher's beliefs about, or understanding of, teaching and learning.

Real change in education has an impact on the interaction between teacher and learner in the classroom. Changes in organisational set-up or in materials will tend to be relatively superficial unless accompanied by change in teachers' behaviour and belief. We might say, then, that changes lower in the list (d. and e.) are more fundamental than those higher up. Change in teacher belief or behaviour is also relatively more difficult to accomplish because it is more personal, because classrooms are private environments and because beliefs are sometimes not outwardly manifest.

The innovative process is a process, however, and one should not expect, therefore, that change at the various levels will occur simultaneously. Understanding and commitment will typically grow in the course of successful implementation. It is quite normal, as Fullan (1989) points out, for behavioural change to precede change in understanding or belief rather than vice versa. Mastery of a new technique may lead into a change of attitude - a point of relevance to in-service teacher education.

The second dimension of innovation is its scale. Innovations vary greatly in how widely they are implemented and in the numbers of people involved. The range may be from a single individual in one institution to an entire national system of education. In the world of private sector ELT, innovation tends to be relatively small scale, involving groups of individuals trying out new ideas in their institution. World Bank sponsored projects, on the other hand, tend to be large scale, involving thousands of people across a whole nation. Implementation processes differ accordingly - with management considerations having greater salience in large scale projects.

There are similarities, however. As Fullan (1989:9) points out, the effectiveness of even quite large change projects

> ......stands or falls with the degree to which front-line implementers (i.e. individual teachers) use new practices with some degree of mastery, commitment, and understanding.

In this paper, we are primarily, but not exclusively, concerned with change on a larger scale than that involving a few individuals in a single institution but on a smaller scale than a national project. Some of what is said may, however, have relevance to change projects at either extremity of the cline. We also need to distinguish between specific innovations and clusters of innovations. The latter are often called reforms. Finally, the focus of change may be a specific innovation (e.g. a new examination) or an enhancement of organisational capacity, or both - though it is usual for one or the other to take priority.

2. Conceptualising change

Before discussing factors affecting implementation, it may be useful to make a number of initial observations on the phenomenon of educational change.

2.1 'Innovation' is a seductive term. Its political economy, however, is such that the likely benefits are often oversold to gain acceptance and resources (Hurst 1983). The reality is that many innovations deliver less than is initially promised. Some turn out to be worthless, and a few are motivated less by an interest in solving problems than by a lust for the social cachet of innovativeness. Faced with claims for innovation, there is, therefore, some justification for caution and scepticism.

Innovation is also sometimes politically motivated and this can mean (i) that too many changes are introduced at once in an ill-coordinated way, and (ii) that changes are introduced prematurely before proper trialling. The result may be that teachers are overloaded. It would be appropriate in this situation to attempt to scale down the scope of change.

2.2 When innovations fail, teachers are often blamed. They are said to resist change. The phrase has a superficial explanatory allure, but is ultimately unproductive. First, it is value loaded in that it assumes the innovation is good and opposition wrong. It thereby delegitimises dissent, which may, of course, be perfectly well-founded either because the idea is not so good after all or because circumstantial factors impede its implementation. More seriously, it is reductive in positing a sort of blind non-rationality on the part of teachers. It seems to pre-empt further enquiry.

As Hurst (1981:185) observes, greater success in implementation will accrue to '......strategies that postulate rational and logical factors'. It will do so because they are better able to uncover the root causes of difficulty and suggest measures for overcoming them. A practical corollary is that if a teacher attending a teacher education course says, 'it wouldn't work in my class', we will have to accept that s/he is probably right - rather than talk self-righteously of resistance.

The wider point is that is that we need to examine change from the 'inside', to adopt a phenomenological perspective that enquires into the meaning the recipient brings to the new information.

2.3 Change usually involves loss, anxiety and risk. There is the risk of a loss of classroom control, and of disapproval from students, peers and authority. There is the 'burden of initial incompetence' (Macdonald and Rudduck 1971) as the teacher abandons the familiar and fumbles with the new. Trying out something new can also bring an uncomfortable exposure. Being in a trial, Shipman notes, often means being on trial. All this means that innovation is demanding. It takes effort and time; time to acquire clarity about what is involved.

2.4 When they first emerge, innovations are seldom fully adapted to their contexts of proposed use. A period of trial, experimentation and adaptation is usually required. The innovating agency should be tolerant of reinterpretation, and of the different versions of the change that emerge from user's adaptations. Success in implementation is not to be measured, then, by degree of compliance but by successful adaptation at 'street level'. Fidelity to original conception is in general negatively related to successful implementation.

3. Factors in the implementation of change

Success in implementing change depends on three categories of factors: the nature of the innovation itself, the transmission of the innovation and the management of change.

Perhaps the most important is the innovation itself. Some viewpoints assign the greatest importance to the transmission process but in so doing they devalue the critical reasoning capacities of the target audience. They claim implicitly that if we communicate the idea effectively, all will be well. But this is not so. The innovation may be rejected on account of its failings. Additionally, we may question the assumption that all innovation is exogenous and therefore stands in need of dissemination.

Another viewpoint that acceptance of change is in some way contingent on the character of the receiving agency, be it an individual or an institution, has led to an unproductive search for characteristics of a psychological or sociological kind that correlate consistently with innovativeness. However, the dependent variable, a stable propensity to accept or reject innovations, is as Hurst (1983:43) suggests, probably mythical. People do not, any more than institutions, conveniently divide into those habitually adopting and those habitually rejecting innovations. A more plausible view is that one and the same individual or institution may be both welcoming of or resistant to change depending on its nature.

Again, we are driven back to the properties of the innovation as determinants of its acceptability. So, a suitable question is - what are the conditions that enhance acceptability?

4. Innovations: conditions for acceptance

Several writers suggest that potential adopters assess innovations according to some cost-benefit calculus. The following are important elements in the calculation.

The change should offer a relative advantage over existing practice, and the probability of the alleged benefits accruing should be high. The change should also be cost efficient; that is, the ratio of benefit to effort should be better than existing practice. Innovations which require consistently more work but offer relatively few gains over existing practice are unlikely to enjoy success.

The change should be perceived as beneficial and feasible in terms of adopter's value systems and working conditions. Innovations are more readily adopted to the extent that they are congruent with existing values and practices. Those which embody unfamiliar values or require a radical reconceptualisation of teaching style have a correspondingly reduced chance of successful diffusion. Macdonald and Rudduck (1971) show how the dissemination of the Humanities Curriculum Project was made more difficult by the unfamiliarity of the teacher's role as neutral chairman of discussion.

A related point is that complexity and ambition can impede successful implementation. Complex innovations are those which require substantial amounts of unlearning-relearning, and ambitious ones are those where the scope of the change is large in relation to the capacity of the receiving system, where large numbers of people are involved, and whose maintenance is time-consuming and elaborate. Ahrens (1991) notes, for example, that one of the causes for the breakdown of the Gujerat Radio INSET project was that '.......the degree of ELT innovation was too big'. The lesson may be that the 'alternative of grandeur' (Havelock and Huberman 1977) should be eschewed in favour of smaller scale, more incremental change.

The risks of change should be reasonable to participants. One way of reducing perceived risk is to allow potential adopters to observe the innovation in use in 'real' classrooms. Huberman (1973) suggests that teachers tend to be more favourable to innovations that they can see put to work in the classroom. Another way is to provide opportunities for trialling the innovation on a limited basis, again in a 'real' classroom.

The innovation should be perceived as practical (Doyle and Ponder 1977). For teachers, this means it should possess the following attributes:

- it should have instrumental content; in other words, describe procedures that have a direct, realistic classroom application. The innovation proposal should not confine itself to rationales or descriptions of abstract principles. It should address 'how to' concerns.
- it should have efficiency, meaning a better yield per unit of effort than existing practice.
- the credentials of the innovation advocates should be credible to teachers. They should be seen as having relevant or comparable experience to the teachers themselves.

Another feature of practicality is that benefits should emerge fairly early in the history of the change project. Teachers, like other people, are in general not good at accepting initial discomfort for deferred benefits.

5. The transmission of innovation

Success in implementation may be influenced by the transmission process; how the idea is communicated to the target audience. There are various models of the transmission process (e.g. Havelock 1969), and an influential typology of strategies for implementing change (Chin and Benne 1969).

We shall not review these here because they have been described elsewhere (e.g.White 1988), and because, although they are useful conceptualisations, they offer few clear guidelines for the practitioner. We can say, however, that there appears to be a basic division of dissemination models into those which see innovation in centre-periphery terms with innovation emerging from a central agency and those which stress the active role of the periphery in initiating innovation.

Among the latter are school-based curriculum innovation movements. Innovations developed at school level largely circumvent the problems of dissemination, and are advantaged in being closer to the point of implementation. This allows for a better fit between the innovation and its context and may encourage a sense of involvement and ownership, which some writers (e.g. Kennedy 1989) stress is important to successful implementation. On the other hand, the assumption that there are sufficient time, resources and expertise at school level to carry out a programme of innovation is often not met in developing countries. Maintaining the existing system is often quite enough of a struggle.

Strategies for implementing change differ in terms of the degree of coercion applied. The most forceful, 'power-coercive strategies' (Chin and Benne 1969), typically involve change imposed from above through, for example, examination reform or ministry circulars. Such methods can produce quick results, particularly in societies accustomed to authoritarian practice. But they are not reliable because they do not guarantee the internalization of the innovation by teachers who, because they work in private settings, may discontinue implementation once external pressure is lifted or distracted elsewhere.

At the other end of the continuum are strategies that coopt teachers into the innovative process and seek to bring about attitudinal change by methods that are almost psychotherapeutic. Whilst these approaches are welcome for their more participatory nature and their attention to the norms that guide practice, they are over-optimistic in their assumption that conflicts of interest can be reconciled. If the target population takes an unfavourable view of the innovation, there may ultimately be little the change agent can do.

Participation is similarly not an unqualified 'good'. Its merits in creating a sense of ownership, in helping to eliminate inappropriate innovations, can hardly be denied. However, it can also be time-consuming and divisive because, as Hurst (1983:19 suggests, it can '....exacerbate and polarise differences of opinion'.

6. In-service teacher training

One of the main vehicles for disseminating educational change is in-service teacher training. The question is not whether this is required but what precise form the training should take.

A conventional form of training is the pre-implementation workshop. Often, this takes place off-site away from the school. It also typically involves a type of instruction that has been labelled 'transmission' (Breen and Candlin 1989). The trainer assumes a missionary role and the trainees for their part are obliged to have faith (ibid.).

This form of training may be quite satisfactory for raising awareness, but for a number of reasons it is of little help in implementing change. First, it is only when teachers actually begin to implement change that they experience the most specific doubts and questions. And it is then that anxiety is at its greatest. This argues the need for continuing support and advice during, as well as prior to, implementation. Otherwise, confidence can quickly evaporate.

Breen and Candlin (1989) also point out that there is typically a gap in thought and action between the workshop and the classroom. Implementation of a new idea is better regarded as a process of trial, evaluation and adaptation which ideally requires extended contact between teacher and trainer. Thus, the 'one shot' workshop with no provision for follow up of attempts at innovation is unlikely to be effective.

There is a further reason for sustained support during implementation. Fullan (1982) argues that major behavioural change requires resocialization, the basis for which is continued interaction over time. Interaction need not only be with experts. Given the finding (Fullan 1982) that teachers often prefer to turn to colleagues rather than external specialists for advice, peers also have an important role. Collegiality, and openness in the classroom, are important assets, then. Both require a climate of trust and support.

In-service-training for innovation is sometimes held to be more effective (Breen and Candlin 1989) if it relates directly to the experiences and problems of teachers in schools, and if it is primarily the teachers themselves that set the agenda for training. It is more effective because (a) it brings the development of the innovation closer to its point of use, relating it more directly to classroom reality, and (b) it helps develop a sense of ownership in relation to the innovation. Both enhance the likelihood of its long term survival and institutionalization.

Training in this view, then, should support the teachers' efforts at innovating in response to their problems, and should be seen as a longer term 'investigative process' (Breen and Candlin 1989:135). The problem is, however, that this long term investigative approach may be prohibitively costly in a situation of economic stringency. It also depends on a degree of teacher confidence and initiative that may not be forthcoming where there is habitual passivity in the face of 'expert authority'. The challenge for the change manager is to evolve modalities of training which respect cost and cultural constraints, which deliver long term support for attempts at innovation, but which avoid the deficiencies of 'transmission training'.

There are implications in this for the location of training for innovation. In- school training is probably preferable. Off-site training removes the participant from the preoccupations of daily life, and this may help concentration where awareness-raising is the objective. But it also means training away from the social reality of the school. A possible consequence is that the enthusiastic but solitary messenger returning from an INSETT course may find it difficult to convince colleagues of the practicality of the new idea, and to persuade them that the risks are worthwhile.

The best form of in-service training for innovation, then, is that which is on-going through the implementation process, that which takes place close to the point of implementation, that which involves demonstration of new practice as well as explanation and feedback on change attempts, that which offers opportunities for practice and trial, and that which comes in a variety of forms: workshops, frequent consultant visits, informal peer conferences.

7. The role of the school principal

Fullan (1989:15) points out that the school or institution is the level of organisation which is closest to the individual implementer, most salient in his daily life, and as such it '....presents the most powerful set of immediate conditions determining the degree of change (or non-change)'.

It is not surprising then that research evidence (Fullan 1982) assigns the school principal an important role in the implementation of change. The active support of the principal is vital, for, in Fullan's words (1982:71), his actions '.....serve to legitimate whether a change is to be taken seriously and to support teachers both psychologically and with resources'.

Support needs to extend beyond verbal endorsement to actions such as securing the assistance of consultants, arranging additional resources where necessary, protecting implementers from excessive demands on their time, and recognising and rewarding implementer efforts. In general, effective change requires a combination of pressure and support, and the school principal may be a source of both.

8. The management of innovation

Hurst (1983) points out that innovative projects have an experimental character in that a considerable period of trial and adaptation is often necessary to achieve a better fit between innovation, user and context. This distinguishes the management of innovation from management as the routine maintenance and administration of existing systems. Different managerial skills are required.

What is needed above all in innovation is the monitoring of implementation. Participants' reactions need to be monitored and procedures established for conveying information from individual teachers to administrators and facilitators. Then corrective action can be taken to overcome inevitable difficulties and disincentives. Hurst (1983) identifies three kinds of corrective action:

(a) better communication to improve participants' understanding of the innovation (e.g. additional in-service assistance)

(b) modification of the innovation to suit users' requirements

(c) assistance for adaptation of the innovation by users (perhaps by adding to users' resources).

If the innovation continues to prove unworkable or unpopular, it should be abandoned as not such a good idea after all.

The emphasis on monitoring contrasts with much common practice where too much attention is given to the initial design and dissemination of change at the expense of implementation. Hurst (1981) points out that it is a profound mistake to think that all difficulties can be foreseen in advance. The inevitability of unforeseen difficulty needs to be accepted. What is important is that there is a swift response to implementation problems, and this implies the retention of contingency reserves of time and resources. It also requires short flexible chains of command.

To participate in innovation is to incur risks. Part of the business of management is to reduce risks and disincentives to acceptable levels. One way of doing so is to provide opportunities to observe and trial the innovation on a limited basis. Pilot projects in the first phase of an innovative programme are a common and effective device for early identification of areas where adaptation is needed and for more realistic estimates of risks. However, successful pilot projects do not guarantee successful dissemination elsewhere. In fact, because they generally receive special attention, they are, to use Crossley's words (1984:84), 'doomed to success'.

The converse of reducing risks is strengthening incentives. Several writers (e.g. Woods 1988, Kennedy 1989, Morrison 1990) point out that incentives are an essential ingredient in programmes of innovation. From the outset participants need incentives to set against the risks, and if motivation falters during implementation, these may need to be strengthened. Woods (1988), among others, suggests that with funded aid projects one kind of incentive could be the offer of scholarships for overseas study.

9. Projects and sustainability

In recent years much large scale innovation in ELT has been implemented through funded aid projects, particularly in developing countries. This often means special project inputs: a project secretariat, project vehicles, overseas consultants, project photocopiers, and so on. The motivation for 'projectisation' is understandable: to establish an enclave against a hostile economic or social environment, and to implement change according to a coherent plan.

This approach has several disadvantages, however. Special inputs may guarantee short term success, but when they are withdrawn, the programme may collapse: the problem of sustainability (British Council 1989,1990).

A second problem is that if the project bypasses regular administrative channels, it may fail through underutilisation to develop their capacity for administration, research, policy analysis, and evaluation (King 1991). In other words, it may fail to develop institutional capacity, a theme of increasing importance on the agenda of many aid agencies. If one accepts that a strong local institutional capacity is supportive of a self-sustaining and independent change programme, then the lesson for the ELT innovator is that he should work as far as possible through existing administrative channels.

Sustainability of innovation may be enhanced in the following ways:

- The scope of the innovation should not be too large for local resources to sustain after the withdrawal of project inputs. Innovation research (Fullan 1982) consistently indicates that very high levels of external support are negatively related to the long term institutionalization of change.
- Local participants need, as Woods (1988) remarks, to be involved in the innovation process, thereby developing a sense of ownership of the innovation.
- There should be incentives to sustain the motivation of local participants and to offset the inevitable risks and losses.
- Support for teachers' efforts at innovation should be scheduled over long time periods.
- Collegiality and teacher support networks should be developed through, for example, teacher newsletters and the construction of teacher resource centres.
- Realistic time horizons for even modest change should be set. Because senior administrators tend to be oriented more to results than implementation, they sometimes underestimate the time needed for the implementation and routinization of innovations. The result may be perfunctory training, hasty decisions, misinterpreted communication, and exhaustion brought on by the effort of coping with unrealistic deadlines on top of routine work. If innovations are to endure, there needs to be an adequate period of settling in when the new idea is routinized; what Hoyle (1972) calls 'refreezing'.

Perhaps the greatest threat to sustainability is overdependence on external resources. We should then be perhaps thinking less in terms of sustaining innovation beyond the life of a project and more in terms of building institutional capacity. As a UNDP report (1992:16) says:

> Few developing countries have the capacity to formulate, plan, implement and manage ....programmes - and to incorporate these programmes into their overall human development efforts. This inadequacy is often perceived as one of the main obstacles to implementing sustainable human development policies and programmes.

Finally, it should not be forgotten that though they are less tractable to management, qualities such as vision and commitment are important in implementing change. Also, perseverance, patience, and attention to detail are essential in implementing the implementation plan (Fullan 1989).

Several accounts of innovation (e.g. Ahrens 1990) acknowledge that the commitment of key personnel significantly influences the likelihood of successful implementation. Commitment, however, implies stability of project leadership, and, thus, where there is a high staff turnover there may be adverse consequences. These may be reduced by relating responsibilities to positions rather than persons, by encouraging a spread of implementation reponsibilities, and by bearing in mind the importance of continuity in evaluating transfer requests. In general, the organisation of the project needs to be robust and flexible enough to cope with inevitable staff changes in a long term project, and with an unpredictable or turbulent wider environment.

10. Conclusion

Implementing change is essentially a practical skill that experience refines. Practice can, however, also be improved by a better conceptualization of the change process. Indeed, Fullan (1982) argues that a sound conceptualization is an important ingredient of managerial expertise along with subject knowledge and interpersonal skills.

The main purpose of this paper has been to contribute to such a conceptualization by drawing on the theoretical literature and available case studies. Accounts of good practice, and of failure, will continue to have useful place in the improvement of change management in ELT. They can extend the experience of the profession and provide a background of shared referents for analytic discussion. And they make it possible for future commentators to distil more sensitive guidelines for implementing change.

References

Ahrens P. 1990a. 'Case study: West Bengal.' in British Council. 1990. Dunford House Seminar Report 1990. London. British Council.

Ahrens P. 1990b. 'Why do projects break down?' in British Council. 1990. Dunford House Seminar Report 1990. London. British Council.

Breen M. and C. Candlin 1989 'The evolution of a teacher training programme.' in Johnson R.K. (ed). 1989. The Second Language Curriculum. Cambridge: CUP.

British Council. 1989 Dunford House Seminar Report 1989. London. British Council.

British Council. 1990 Dunford House Seminar Report 1990. London. British Council.

British Council. 1991 Dunford House Seminar Report 1991. London. British Council. 1991.

Chin R. and K. Benne 1970. 'General strategies for effecting change in human systems.' in Bennis W., K. Benne and R. Chin (eds.) 1970 The Planning of Change. 2nd edition. London: Holt, Rinehart and Winston.

Crossley C. 1984. 'Strategies for curriculum change and the question of international transfer.' Journal of Curriculum Studies 16, 1.

Doyle W. and G. Ponder 1977. 'The practicality ethic in teacher decision making.' Interchange 8, 3, 1-12.

Fullan M. 1982. The Meaning of Educational Change. Toronto: OISE Press.

Fullan M. 1989. Implementing Educational Change. PHREE Background Series Paper, Document Number: PHREE/89/18. Washington DC: Education and Employment Division, Population and Human Resources Department, World Bank.

Havelock R. 1969. Planning for Innovation: a Comparative Study of Literature on the Dissemination and Utilization of Scientific Knowledge. Ann Arbor, Michigan: Centre for Research on Utilization of Scientific Knowledge.

Havelock R. and A. Huberman. 1977 Solving Educational Problems: the Theory and Reality of Innovation in Developing Countries. Paris: UNESCO, IBE Study.

Hoyle E. 1972. 'The Problems of Educational Innovation 1.' Educational Studies: A Second Level Course: Milton Keynes: The Open University Press.

Huberman A. 1973. Understanding Change in Education: an Introduction. Paris: UNESCO, IBE Study.

Hurst P. 1981. 'Some issues in improving the quality of education.' Comparative Education 17, 2, 185-193.

Hurst P. 1983. Implementing Educational Change: a Critical Review of the Literature. EDC Occasional Papers no 5. London: University of London Institute of Education.

Kennedy C. 1988. 'Evaluation of the management of change in ELT Projects.' Applied Linguistics 9, 4, 329-342.

King K. 1991. 'Capacity-Building for Educational Reseach, Planning and Analysis.' Draft Paper for World Bank, Department of Education, University of Edinburgh.

Macdonald B and J. Rudduck 1971. 'Curriculum research and development projects: barriers to success.' British Journal of Educational Psychology 41, 148-154.

Morrison J. 1990. 'What are the factors that contribute to sustainability?' Dunford House Seminar Report 1990. British Council, London.

United Nations Development Program. 1992. Human Development Report. Oxford: OUP.

White R. 1988. The ELT Curriculum. Oxford: Blackwell.

Woods P. 1988. Pulling out of a project: twelve tips for project planners. ELT Journal 42/3 1988.

FUNCTIONAL CONTROLLED WRITING

Ardeshir Geranpayeh (DAL)

Abstract

This study reports the results of two experiments involving a focus on the rhetorical functions of generalisation and classification in the teaching of the writing skill to EFL learners in Iran. The main research question is whether the teaching of language functions is relevant in the development of writing ability in university EFL learners. The results suggest that teaching language functions has a positive effect on the development of this ability.

1. Introduction

1.1 Introduction

English language teaching (ELT) has gone through various stages of development during the last three decades culminating in the emergence of the communicative approach. Much research has been on the effectiveness of adopting a functional approach in the teaching of English as a second language (TESL). Whether functional teaching is feasible in EFL settings has hardly ever been questioned.

In Iran, for example, the official figures speak of about 8 million learners currently learning English at different levels and institutions throughout the country. The rising fever for learning English does not seem to have affected the older generation instructors' methods. The usual practice is that of structural and grammar-translation methods. There is a tendency amongst the adherents of these methods to resist any change in the teaching curriculum. As Hashemi points out,

> even those who are aware of the pitfalls of the generation-old methods and wish to be innovative in their instruction play lip service to the current teaching trends and continue with the inveterate procedures with which they feel secure. (1992:1)

On the other hand, a needs analysis of the learners indicates that English is mainly used, firstly, as a means of reading academic texts and secondly, as a means of expressing learners' ideas through written discourse to English speakers worldwide. These methods seem to serve the learners' needs so that there is hardly any room for the application of any new method. This study, however, is an attempt to investigate the feasibility of applying a functional approach to the teaching of the writing skill in Iran.

The paper begins with a discussion of the "usage-use" dichotomy and its application to the teaching of the writing skill. Then the method of adopting a functional approach to teaching writing to EFL learners is explained. The results of two experiments are reported and analysed in this regard. Finally, conclusions are drawn and theoretical as well as pedagogical implications are proposed.

1.2 Background

1.2.1 Usage-use dichotomy

It was Widdowson (1978) who first made a distinction between usage and use. He argues that the usage-use distinction is related to Chomsky's competence-performance distinction. Competence 'refers to a person's knowledge of his language, the system of rules which he has mastered so that he is able to produce and understand an indefinite number of sentences' (Crystal 1991). Competence, in Chomskyan terms, is an idealized conception of language as opposed to the concept of performance which is a set of specific utterances produced by native speakers. In other words, competence is the language user's knowledge of abstract linguistic rules; when this knowledge is put into practice, it is called performance. According to Widdowson (ibid.), usage and use are two aspects of performance. Usage is that aspect of performance 'which makes evident the extent to which the language user demonstrates his knowledge of linguistic rules' (ibid:3). Furthermore, Widdowson clarifies the issue when he says:

> use is another aspect of performance: that which makes evident the extent to which the language user demonstrates his ability to use his knowledge of linguistic rules for effective communication. (ibid.)

In short, we see examples of usage in grammar books when knowledge of competence is realized through the citation of sentences which illustrate the rules. Such sentences only reveal the language user's ability to use his linguistic knowledge of rules without any communicative purpose. Instances of use are the result of that knowledge being put into practice for effective communication.

1.2.2 Approaches to the teaching of the writing skill

Having explained the usage-use dichotomy, we now need to discuss the way(s) it can be utilized in the teaching of the writing skill. For the purpose of this study, teaching the writing skill is viewed from two perspectives: traditional and modern. The former is based on usage while the latter is based on use.

1.2.2.1 Traditional: focus on the composing skill

The traditional view usually focuses on the composing skill of the learners. By this, I mean the learners are required to compose grammatical sentences irrespective of their functions within a piece of discourse. Widdowson (1978) holds the same view when he reviews the traditional grammar exercises done in classrooms and concludes that:

> as long as they aim at providing practice in correct sentence construction they are directed at the development of the composing skill without regard to the part this skill plays in the writing ability.(ibid:115)

This is due to the nature of such exercises which concentrate on separate sentences in isolation from a context. Such exercises lack the character of instances of use because they do not have any communicative purpose. Rather, they are intended to reveal the learner's knowledge of the language system and the ways it is manifested. They are, in other words, exercises in usage.

Then, the question arises as to whether there is any way to introduce a use orientation in the teaching of the writing skill and, if so, whether it is possible to direct usage exercises towards the development of the writing ability.

1.2.2.2 Modern: focus on the writing ability

The development of the writing ability is the main focus of modern approaches to the teaching of the writing skill (see Arnaudet and Barrett 1984; Kaplan, Robert, and Shaw 1983; Mckay and Rosenthal 1980; Raimes 1991; Reid and Lindstorm 1985; Trimble 1985; Zamel 1987). These approaches are mainly based on use orientation. To have a use orientation one has to devise one's exercises in such a way that they aim at developing natural language behaviour. To achieve this goal, Widdowson proposes two kinds of exercises: preparation and exploitation. By the former, he means exercises which precede a reading passage and force the learner to participate in actual writing; by the latter, he means exercises which follow the reading passage and exploit it for the purpose of practice material.

In preparation exercises, the instructor chooses a reading passage which realizes a certain function, e.g. classification. The exercises that precede this reading passage are all directed to the production of a passage similar to the one to be followed. Hence, the comprehension of the new passage becomes easier. Moreover, since the preceding exercises are all based upon composing activities in which students compose sentences, the composing ability of the learners improves. These exercises are different from those of the traditional approach in that they follow a process of gradual approximation.

What is meant by gradual approximation? It is a general strategy exposed to the learner to develop his communicative abilities in the foreign language. According to Widdowson, gradual approximation:

> begins by providing exercises within the scope of the learner's (limited) linguistic competence in English and then gradually realizes its communicative potential by making appeal to the other kinds of knowledge that the learner has.(1979:76-7)

This strategy involves the learner both in usage and use activities, in which the starting point is the sentence and the target is discourse. Therefore, in our language learning pedagogy activities in both usage and use are required. However,the suggestion is that the main orientation should be toward using language as communication. That is, use activities play an important role in pedagogy. The strategy that bridges the gap between usage and use is, then, called gradual approximation. (see Widdowson, 1979: 75-85)

The second kind of use exercises is exploitation. These exercises follow the reading passage and use it for the purpose of practice material. They should exploit the contextualization provided in the reading passage and should 'use the passage as a basis for the development of the writing ability. '(Widdowson 1978:123). As in preparation exercises, the practice of particular aspects of grammar can be associated with the writing of simple instances of discourse by means of gradual approximation, thus giving it a free orientation. It should, nevertheless, be noticed that the end product is not going to be a [illegible] version of the reading passage any more. Rather, it is going to be an [illegible] which is somehow theoretically related to it e.g. a summary of the passage.

[illegible] approach [illegible] exploitation exercises through different [illegible] discourse is a common approach. The process [illegible] preparation exercises. Various manipulations [illegible] could a simple account based on the example [illegible] description used in the passage. Such a process [illegible] that aspect of usage found in the passage. Since the procedure is controlled, it can result in the development of more elaborate versions with different degrees of complexity according to the learner's competence. At the end of the exploitation exercises the learner should be able to write pieces of discourse (paragraphs) functionally similar to those used in the reading passage. Hence, the learner can communicate effectively in writing within the limited scope he has been exposed to. The scope can be enlarged according to the needs of the learners (see ibid: 126-134).

Then comes a [illegible] the procedure outlined above in the development of [illegible] EFL learner. This is the focus of the [illegible] research.

1.3 Statement of purpose

The main research question of this study can be stated as: "Is it effective to teach language functions through controlled writing?" The hypothesis is that teaching language functions [illegible] instances of use through controlled writing should affect the writing ability of students, in this case EFL learners. This is in accordance with the current trend in the teaching of the writing skill in which use is emphasized. As mentioned earlier, exercises in use involve the learners in actual communication. What is meant by [illegible] in writing. It is a process of non-verbal transfer of information [illegible], for example, are non-verbal carriers of information which the writer [illegible] to communicate ideas to his readers. Since paragraphs can serve different functions, [illegible] seem to be necessary parts in any language teaching syllabus based on the characteristics of use.

In brief, language functions resemble the major thought patterns governing a piece of discourse. If a learner internalizes these functions, it is assumed that he should be able to perform them effectively. In the case of writing paragraphs, internalizing the language functions should result in the improvement of the writing ability. The degree of improvement, however, is a matter that this research intends to investigate. Other research questions that were posed are as follows:

1- Is there a hierarchy of difficulty among the functions?

2- If there is a hierarchy of difficulty among the functions, does it affect the learning process(es) of the functions?

The alpha level for all statistical decisions was set at 0.01.

BEST COPY AVAILABLE

2. Method

2.1 Subjects

Two groups of subjects participated in these experiments, the experimental group and the control group. Treatment was intended to be applied to the experimental group only.

The experimental group (Group A) consisted of 38 first-year Iranian English majors studying at Azad University in Tehran. They all participated in the course "Grammar and Writing II", in which the students were taught to combine simple sentences and construct compound-complex sentences to form paragraphs. The course was based on a usage-oriented method of the kind described earlier. The control group (Group B) consisted of 41 first-year Iranian English majors studying at Shahid Chamran University in Ahwaz. They were also taking the course "Grammar and Writing II" with the same syllabus and method as those of Group A subjects.

In order to assess the subjects' language proficiency, a version of the 100-Multiple Choice Nelson Quickcheck Test - the reliability of which had been reckoned to be 0.93 - was given to both groups. Only those who scored over 50 were selected to take part in the experiments. Thus, 31 and 23 subjects participated in Group A and Group B, respectively. The claim that the groups were homogeneous is based on a t-test conducted to compare the scores of the two groups. The t observed was $t_L = 2.5476$, d.f. = 52, $p < 0.01$. This allows us to infer that the groups were homogeneous.

2.2 Materials

The preparation of materials was a very difficult and critical task in these experiments. Two factors had to be taken care of: 1) the relevance of materials to the level and fields of the subjects, and 2) the utilization of preparation and exploitation exercises preceding and following a reading passage for the development of functional writing. Prior to decisions about materials and activities, the functions themselves had to be determined. Language functions can be classified in various ways. However, for the purpose of this research, they are divided into the thematic and the supporting functions. The former, sometimes called rhetorical, refer to very broad functions like generalisation, elaboration, and classification. They usually represent the main propositional development in a piece of discourse.The latter functions deal with supporting acts that link the smaller units of information in a piece of text,i.e., clarification or exemplification (see Widdowson 1978: chapter 5). Supporting functions are,elsewhere (Trimble 1985), equated with rhetorical techniques. They are devices a writer uses to relate the units of information in a paragraph to one another and to relate the paragraphs of a discourse to each other. Moreover, they are termed techniques, for it is not common to find a whole paragraph comprising them.

The two rhetorical functions of generalization and classification, which have wide potential use in most academic areas, were selected. The function of generalization was used in Experiment I and that of classification in Experiment II. The advantages of selecting two functions are as follows. Firstly, a particular function might, for various reasons, be perceived with difficulty by the learner, and such difficulty or simplicity may influence the results. Secondly, there might be a hierarchy of difficulty in learning functions, i.e. one function may be learned better than the other. The selection of the second function will clear the ground.

To control the teaching situation of the two experiments, it was decided to focus only on one rhetorical technique (one pattern) for the development of both generalization and classification. This led to the selection of exemplification as the appropriate rhetorical technique for the development of the functions.

Having determined the functions, some passages featuring generalisation and classification supported by exemplification were adapted and (carefully) organized to fulfil the requirements of the preparation and exploitation exercises. The researcher prepared two syllabuses for the treatment in the experiments. The syllabuses consisted of two main parts, namely Presentation and Exploitation. Both the teacher and the students were provided with the syllabuses during the treatment.

2.3 Procedure

2.3.1 Treatment

In Experiment I the subjects in Group A were required to attend two 90-minute sessions on two successive days. The presentation section consisted of two parts: (a) making the subjects familiar with the statements of general and specific information, and (b) checking their comprehension with questions regarding general and specific points. Then, the subjects were given a reading passage followed by exploitation exercises. These exercises required the subjects to find general and specific terms or to complete diagrams representing the outline of the passage. The same procedure was applied to some other passages within the syllabus. Some of the activities required the learners to form sentences from jigsaw words. Then, the subjects were supposed to join the sentences thus formed to complete a paragraph. The role that each sentence played within the paragraph was underlined and the subjects were allowed to add any transitional marker (rhetorical signal) that established the coherence of the paragraph.

In Experiment II, the subjects in Group A were taught the function of classification. The same procedures as those of Experiment I were adopted. This time, however, the emphasis was on the formation of categories (classifications) and their members, on the basis of their shared features. The treatment was then complete.

2.3.2 Evaluation

Based on the results of an earlier pilot study conducted on other subjects, it was decided to have the subjects write a controlled paragraph for each experiment. The paragraphs were divided into 5/6 single statements. Each statement had some missing words followed by a parenthesis which included some of the words needed to complete the statements. Moreover, the function/technique resembling each statement was underlined (see Appendices A and B). The subjects were required to complete the sentences and to add the necessary words that established the coherence of the paragraphs. Finally, they had to write their paragraphs on separate sheets of paper. These tests were distributed among both the experimental and the control groups. The paragraphs thus formed were then collected for evaluation.

The next task was to choose a criterion for evaluation. The existing criteria for evaluating the writing skill (discussed by Jacobs et al. 1981) were neither objective nor adequate for this research. Therefore, it was decided to devise a new criterion of assessment based on three factors: the function that each sentence played within the paragraph (F); the rhetorical signal used (R) if it were necessary; and the grammaticality of the sentences (G). The following grading system was employed for each sentence: 2 points to each sentence if the desired function /technique was fully performed; 1 point if the proper rhetorical signal was used; and 1 point if the grammaticality of the sentence was observed. A negative point was given to those sentences violating the normal form,i.e. the subject did not use all the words provided in parentheses for writing the sentences. Table 1 shows how each paragraph was scored.

TABLE 1
Evaluation Format

	SENTENCE						
	1	2	3	4	5	6	Σ
F	2	2	2	2	2	2	12
R	* -	1	1	* -	1	* -	3
G	1	1	1	1	1	1	6
V	@ -	@ -	@ -	@ -	@ -	@ -	@ -
Σ	3	4	4	3	4	3	21

F= Function
R= Rhetorical Signal
G= Grammaticality
* No rhetorical signal required
@ No violation in constructing the sentence
V= Violation (negative score)
Σ= Total Score

The paragraphs were then numbered and shuffled. The paragraphs were rated by a second judge in addition to the researcher. Later, the results of the two scores were compared. The degrees of correlation between the two raters' scores for the experimental group were r1=0.86 and r2= 0.88 in Experiment I and Experiment II, respectively, and those for the control group were r3= 0.93 and r4= 0.93 in the experiments, respectively. The high correlations thus confirmed the objectivity of the evaluation procedure.

3. Results

3.1 Experiment I

To find out whether the difference between the writing scores of the two groups was significant, a t-test was performed. The t-test revealed that the experimental group scored significantly higher than the control group, t2= 8.4421, d.f.= 47, p<0.01. Other statistical results obtained are given in table 2.

TABLE 2
Mean Scores of the Subjects on Each Individual
Factor Affecting the Total Writing Score:
Experiment I

	Group A						Group B				
	F	R	G	V	Σ		F	R	G	V	Σ
Mean	9.1	2.1	2.5	.44	13.4	X	3.7	.33	.71	.4	4.38
%	91	70	50	44	74	%	37	11	14	38	24

F= Function
R= Rhetorical Signal
G= Grammaticality
V= Violation
Σ= Total Mean
X= Mean score

3.2 Experiment II

Another t-test was performed on the writing scores of the two groups of subjects in Experiment II. The t-test, once again, revealed that the difference between the two mean scores of the two groups was significant,t3= 5.023, d.f.= 40, p<.01. Table 3 illustrates the average scores of the subjects on each individual factor affecting the total writing score in the experiment.

TABLE 3
Mean Scores of the Subjects on Each Individual Factor Affecting the Total Writing Score: Experiment II

	Group A						Group B				
	F	R	G	V	Σ		F	R	G	V	Σ
Mean	10.8	1.2	4	.6	15.4	X	7.9	.04	2.2	.7	9.4
%	89.6	41	81	62	73	%	66	1.3	36	73	[illegible]

F= Function
R= Rhetorical Signal
G= Grammaticality
V= Violation
Σ= Total Mean
X= Mean score

3.3 Experiment I VS Experiment II

Table 4 contrasts table 2 of the first experiment with table 3 of the second experiment.

TABLE 4
Contrasting the Details of Mean Scores of the Factors Affecting the Total Writing Score: Experiment I VS Experiment II

	Group A						Group B				
	F	R	G	V	Σ		F	R	G	V	Σ
EXP I	9.1	2.1	2.5	.44	13.4	X	3.7	.33	.71	.4	4.38
	91	70	50	44	74	%	37	11	14	38	24
EXP II	10.8	1.2	4	.6	15.4	X	7.9	.04	2.2	.7	9.4
	89.6	41	81	62	73	%	66	1.3	36	73	45

F= Function
R= Rhetorical Signal
G= Grammaticality
V= Violation
Σ= Total Mean
X= Mean Score

It reveals the following. Firstly, Group B subjects scored significantly better in Experiment II than in Experiment I, $t_4 = 3.9149$, d.f.= 40, $p<0.01$. Secondly, there was no meaningful difference between the mean scores of the two experiments for Group A subjects, $t_5 = 2.3358$, d.f.=47, $p<0.01$. Thirdly, the rhetorical signals used for the function of classification were more difficult than those used for that of generalization. Group A scored 70% and 41% while Group B scored 11% and 1.3% in Experiments I and II, respectively. Finally, the grammaticality of the sentences in Experiment II had improved when compared to that of Experiment I: 50% and 81% for Group A, and 14% and 36% for Group B in Experiments I and II, respectively.

4. Discussion

The t-tests (t2 and t3) performed in these experiments confirmed the idea that there was a meaningful difference between the mean scores of the two groups after the treatment. Group A scored significantly higher than Group B in both experiments. Since there was no significant difference between the two groups (t1) before the experiments, this would seem to suggest that the improvement in the scores of the experimental group was due to the treatment they had received. That means the teaching of the functions appears to have affected the writing ability of the learners.

A word of caution. This research was conducted on intact groups. Every effort was made to control possible extraneous factors which might have affected the results. But like all other research of this kind, it has its limitations. The great difference between the writing abilities of the two groups after the experiments might cast doubt on the results and suggest that perhaps one group was somehow severely disadvantaged. But this was not the case. Applying treatment to only one group may disadvantage the other group in any experimental design. Giving the advantage of treatment to the experimental group is the procedure normally adopted in experimental designs. The purpose is to test whether the 'advantage' is really an advantage, leading to meaningful changes in the performance of the subjects. If it does change the performance, which in the case of this research it did, it is usually interpreted that perhaps it was due to the treatment effect the subjects had received. It was argued in section 1 that practice in usage, unless accompanied by practice in use, does not automatically yield instances of use. In developing writing ability, mere exercises in composing, which both groups were exposed to, does not necessarily lead to the development of writing ability. Returning to the present research, only the experimental group who were exposed to use activities, in addition to usage ones, were able to perform significantly better in writing. This, furthermore, may indicate the reason for the great difference between the two groups; perhaps the control group, because they were exposed only to composing exercises, could not develop their writing ability during the time limit of this research. In either case the importance of use activities in the development of writing ability cannot be ignored.

Moreover, one may conclude from t4 that the function of classification was an easier task than that of generalization for Group B subjects, that is to say, that there is a hierarchy of difficulty between the functions. However, t5 illustrates that, even if there were such a hierarchy, it did not affect statistically the learning of the language functions by the experimental group. Group A performed equally well on both functions. This may lead us to the conclusion that in learning the language functions, the simplicity or the difficulty of those functions does not seem to play a significant role.

Finally, the results reveal that the rhetorical signals used in Experiment II were more difficult than those in Experiment I for all subjects. It can be seen that the rhetorical signals employed in Experiment I are **for example, for instance,** and **such as**. These signals all indicate that an exemplification is to follow. The rhetorical signals required in Experiment II are **first, second,** and **third.** Although these can also be considered signals of exemplification, they indicate that a kind of enumeration is going to take place. These rhetorical signals, according to Halliday and Hasan (1976), are among 'temporal conjunctions'. They differ from all other rhetorical signals in that they do 'occur in a CORRELATIVE form, with a cataphoric time expression in one sentence anticipating the anaphoric one that is to follow'(1976: 263). That is, once a learner uses the rhetorical signal **first**, he is inclined to use **second,** etc. On the other hand, if he misses the first signal, it will be difficult for him to anticipate the next. Returning to the present study, the difficulty of temporal conjunctions for the subjects might be due to the fact that once they missed the first signal, i.e. **first**, they could not anticipate the next. This may be why it appeared in the results that perhaps the rhetorical signals of Experiment II were more difficult than those of Experiment I.

5. Conclusions

We have argued in this paper that, in spite of the large amount of research into communicative teaching, little attention has been given to the feasibility of the approach in the teaching of writing in EFL settings. To test whether this method of teaching is applicable to teaching writing to EFL learners, two experiments were conducted. The experiments were carried out to determine the effectiveness of functional teaching in the development of the writing ability of EFL learners. The results suggested that the method was effective and feasible if a process of controlled writing was intended.

The study can contribute to writing research in two respects, theoretical and pedagogical. As far as theoretical implications are concerned, the following conclusion is plausible. Language functions play an important role in the development of the writing ability of EFL learners. The results of this research showed that those learners who had acquired the language functions could perform better in their writing task.

This study has pedagogical as well as theoretical implications. Practitioners can take insights from this research for their classroom activities. The findings of the present study will benefit those EFL teachers willing to adopt a communicative approach to the teaching of the writing skill. They can use the same procedure adopted here: the process of gradual approximation. Finally, it was also found that controlled writing appeared to be an appropriate means for teaching the functions of the English language. This method of teaching the language functions involves the learners in the process of gradual approximation.

Acknowledgements

I am grateful to Alan Davies, Margaret Murphy, and an anonymous EWPAL reviewer for their insightful comments and suggestions on an earlier draft of this paper.

References

Arnaudet M.L. and M.E. Barrett.1984. Approaches to Academic Reading and Writing. Englewood Cliffs, N.J.: Prentice-Hall.

Crystal D. 1991. A Dictionary of Linguistics and Phonetics. Oxford: Blackwell.

Halliday M.A.K. and R. Hasan. 1976. Cohesion in English. London: Longman.

Hashemi Mohammad Reza. 1992. 'Feasibility of functional approach to EFL teaching in Iran'. Unpublished paper presented at TESOL '92 Convention, Vancouver.

Jacobs H.L., S.A. Zinkgraf, D.R. Wormuth, V.F. Hartfiel and J.B. Hughey. 1981. Testing ESL Composition. Rowley, Mass: Newbury House.

Kaplan R.B. and P.A. Shaw. 1983. Exploring Academic Discourse. Rowley, Mass: Newbury House.

Mckay S. and L. Rosenthal. 1980. Writing for a Specific Purpose. Englewood Cliffs,N.J.: Prentice-Hall.

Raimes A. 1991. 'Out of the woods: emerging traditions in the teaching of writing'. TESOL Quarterly 25/3: 407-30.

Reid J.M. and M. Lindstrom. 1985. The Process of Paragraph Writing. Englewood Cliffs, N.J.: Prentice-Hall.

Trimble L. 1985. English for Science and Technology: a Discourse Approach. Cambridge: Cambridge University Press.

Widdowson H.G. 1978. Teaching Language as Communication. Oxford: Oxford University Press.

Widdowson H.G. 1979. Explorations in Applied Linguistics (Vol.1). Oxford: Oxford University Press.

Zamel V. 1987. 'Recent research on writing pedagogy'. TESOL Quarterly 21: 697-715.

Appendix A

Writing test used in experiment I

Instructions: Construct a paragraph based on the incomplete sentences using the words in parentheses. Add necessary words that establish the coherence of the paragraph. The role that each sentence plays within the paragraph is underlined. Write your paragraph on a separate sheet of paper.

Words and Meanings

-- ..... most interesting ..... actual names ..... which ..... associated ..... meanings of the words.

(some, words, English, are, people)

GENERALISATION

-- ..... boycott, ... , ... the case of ..... who ..... by his .....

(word, derive, Sir Charles Boycott, ostracised, tenants)

EXEMPLIFICATION OF WORDS

-- ..... levi's; these popular ..... Levi Strauss who .....

(is, blue jeans, named, after, first, manufacturer, jeans)

ANOTHER EXEMPLIFICATION

-- Perhaps the most ..... sandwich, named for ..... who .....

(is, Fourth Earl of Sandwich, created, quick, portable, meal)

ANOTHER EXEMPLIFICATION

-- ..... this unique category....

(words, include, lynch, watt, davenport, zeppelin)

CONCLUDING SENTENCE: FURTHER EXAMPLES

Appendix B

Writing test used in experiment II

Instructions: *Construct a paragraph based on the incomplete sentences using the words in parentheses. Add necessary words that establish the coherence of the paragraph. The role that each sentence plays within the paragraph is underlined. Write your paragraph on a separate sheet of paper.*

Your Library

-- There are ..... kinds ..... be found .....

(three, basic, materials, can, good, library)

TOPIC SENTENCE: CLASSIFICATION

-- ..... on all subjects, ... languages.

(are, books, both, English, other)

FIRST MEMBER EXEMPLIFIED

-- These books ..... according ..... in a ..... called.....

(are, organize, subject, title, author, central, file, card, catalogue)

FURTHER EXEMPLIFICATION OF THE FIRST MEMBER

--... there are.... which include .... and which ....be used..

(reference, works, encyclopedias, bibliographies, dictionaries, must, library)

EXEMPLIFICATION OF THE SECOND MEMBER

-- ..... there are .....__ which are ..... in racks.

(periodicals, magazines, newspapers, pamphlets, filed, alphabetically)

EXEMPLIFICATION OF THE THIRD MEMBER

-- Like ..... , periodicals cannot .....

(reference, works, removed, library)

EXPRESSING A COMMON FEATURE BETWEEN MEMBERS TWO AND THREE

DEVELOPMENT AND VALIDATION OF A TRANSLATION TEST

Behzad Ghonsooly (DAL)

Abstract

Translation testing methodology has been criticized for its subjective character. No real strides have so far been made in developing an objective translation test. In this paper certain detailed procedures including various phases of pretesting have been performed to achieve objectivity and scorability in translation testing methodology. In validating the newly-developed objective translation test, the following research questions are asked: a) What is the reliability of scores of the translation test and how does it compare with the criterion measure?, b) What is the concurrent validity of the test and of the criterion measure?, c) Are there any factors such as underlying constructs that the translation test and each subtest of the criterion measure may assess? The following general hypothesis is proposed: in measuring the English proficiency of Iranian EST university learners, a translation test is as valid and reliable as a standardized objective test. Results showed significant reliability for the new test.

1. Introduction

As early as the beginning of the twentieth century,the grammar-translation method was disfavoured on the grounds that it did not take into account speaking, writing and listening as important skills of second/foreign language teaching and learning. It was, therefore, excluded from the teaching paradigm. With the exclusion of the traditional method, translation as a testing device was excluded too. Lado(1964) argued that translation tests were highly subjective, referring to the interference of the teacher's taste in scoring a translation test, which resulted in its unreliability. It was also maintained that translation tests lacked the property of scorability (Lado 1964; Harris 1969). The scorability of a language test is defined in terms of how well and easily it is scored. This idea of scorability, which has served as one of the distinguishing features between essay or subjective type questions and the so-called objective tests, draws upon the notion of convenience and speed in scoring a test. Thus, a well-designed test which collects all the responses on a separate sheet and can be scored by machine is much more convenient and less time-consuming and thus more scorable than one which has the responses scattered in the pages of the test. In fact, one might just imagine how difficult an undertaking it may appear for a teacher who is to correct an average number of, for example, 40 students' responses on a rendered text with a length of one or in some cases more than one paragraph.

Taking this into account, it has been argued that scoring essay-type questions including translation tests is not as easy and convenient as,for instance, a multiple-choice question; therefore, they have been judged to be too burdensome and time-consuming.

However, attempts have recently been made to revive translation as a useful device for the purpose of language teaching (Titford 1983; Tudor 1987). As a result of this movement to re-assess the potential contribution which translation can make to ELT after Lado's rather sweeping dismissal of it, new theories of translation have evolved to pave the way for the development of translation teaching activities (see Newmark 1981; Nida 1982). Nevertheless, while translation methodology has been influenced by improvements in translation theory, its testing counterpart has remained untouched. No real advance has so far been made towards constructing an objective translation test to remedy for the above-mentioned deficiencies. This paper is oriented towards the essential procedures for the development of an objective translation test which may fulfil the scorability criterion of the newly developed test and guarantee its objectivity.

2. Design of the study

2.1 Hypothesis and research question

To determine the statistical characteristics of the new translation test, the following hypothesis was adopted: in measuring the general English proficiency of Iranian English for Science and Technology (EST) learners, a translation test would be as valid and reliable as a standardized objective proficiency test. To provide data for testing the hypothesis the following research questions were addressed: a) What is the reliability of the translation test and how does the test compare with the Michigan EFL test? b) What is the concurrent validity of the new translation test and of the criterion measures? c) Are there any common factors such as underlying constructs that the translation test and each subtest of the criterion measure may be assessing?

2.2 Subjects

The total sample of subjects who were exposed to various phases of pre- and post-testing were 315 male and female university students from the Department of Electronics of Tehran University (TU) and Science and Technology University (STU) who had passed ESP courses in the current Iranian educational system. They were supposed to have acquired general English proficiency.

2.3 Instrumentation

Two classes of multiple-choice item tests were administered in this study: the new translation test, which consisted of twenty multiple-choice items and the Michigan test (used as the criterion measure) which comprised forty grammar M/C questions and forty vocabulary M/C questions together with two reading comprehension passages, each of which consisted of five M/C questions.

2.4 Methods of data collection

The decision as to what translation elements should be selected for the construction of the translation test was one of the difficulties in the investigation. Since the content of the translation test was hypothesized to be independent of the content of the materials used in a particular course of instruction, it was not felt necessary to impose any limitation on the content of the test except that the content had to be compatible with the examinees' field of study, namely electronics. Consequently, scientific and technical English texts were chosen as content elements of the translation test. Since each English scientific text (EST) unit of discourse is a coherent paragraph comprising a number of sentences and is too long to be included in the translation test, it was decided to narrow down the task of selection and search for smaller units of discourse, typically sentences. But due to the typological variety of sentences in English, the decision as to which sentence type should be selected posed another problem. It was decided to deal with those rhetorical functions which, as Trimble (1985) argues, are fundamental elements in the organization of an EST paragraph.

2.4.1 Selecting the rhetorical functions

Determining rhetorical functions with regard to the kind and amount of information each provides the reader with, Trimble (1985) distinguishes five major functions and fifteen related sub-functions. Making full use of the rhetorical functions and their related sub-functions in the translation test seemed to be impractical if not impossible. Therefore, setting some criteria for the selection of functions became necessary. Functions and sub-functions were used in the construction of the translation test only if they met these criteria:

1. is always used in written EST discourse;
2. has high frequency of occurrence and usage in academic settings;
3. does not overlap with other functions or sub-functions.

On the basis of the above criteria, the following rhetorical functions and sub-functions were selected

Rhetorical Function	1	Description
sub-function	1.1	physical
sub-function	1.2	function
sub-function	1.3	process description
Rhetorical Function	2	Definition
sub-function	2.1	formal
sub-function	2.2	semi-formal
Rhetorical Function	3	Classification
sub-function	3.1	complete

Rhetorical Function	4	Instruction
sub-function	4.1	direct
sub-function	4.2	indirect
sub-function	4.3	instructional information
Rhetorical Function	5	Visual-verbal relationship

All the examples of the above-selected rhetorical functions used were taken from EST paragraphs. A preliminary version of the test based on the selected rhetorical functions within EST paragraphs was prepared for different phases of pretesting.

2.4.2 Pretesting

One of the fundamental purposes of pretesting is to draw out a variety of responses which can be used as distractors for the final test items. For this reason care was taken over the different phases of pretesting. These are briefly explained here.

2.4.2.1 Phase 1. Pretest with sample population of students

In this phase, one hundred students at TU were pretested. They were both male and female and were randomly selected from 825 Engineering students who had been registered for English proficiency tests such as TOEFL and the Michigan test. These tests are occasionally administered at TU for those students who are eager to get an objective view of their English proficiency. The purpose of this phase was to elicit different alternatives. Hence, a preliminary version of the test, consisting of forty items in an open-ended form, was given to the subjects. They were required to read each EST paragraph and translate the underlined rhetorical function of each paragraph.

2.4.2.2 Phase 2. Pretest with translation expert

The same forty items in an open-ended form were given to two translation experts who were required to write the most desirable translation for each underlined rhetorical function. The purpose of this phase was to obtain the most appropriate response for each item by comparing students' responses for the construction of the test items and to ensure its objectivity.

2.4.2.3 Selecting the alternatives

As to the correct response, only those responses agreed upon by the translation experts were inserted in the tests as the most desirable choices. Other distractors were selected from among students' responses which did not conform to those of the translation experts. But the decision as to what distractors should be selected for each item appeared to be a problem. To solve the unwanted obstacle and to be objective, a tentative criterion was proposed. The criterion was set such that the distractors should have a high frequency of occurrence and be often used by the students. The most common mistakes elicited from students' responses were mainly those of comprehension of the functions, word for word translation and deviant translation including errors of style, grammar and lexicon. Each item was, therefore, given the following arrangement of choices: 1. the correct response, 2. reading comprehension distractor, 3. word for word translation, 4. deviant response distractor.

2.4.2.4 Phase 3. Pretest with sample population of students

After developing the test in M/C form, in order to ensure the difficulty level of the test items, the items were administered to another population of 55 students of Electronics at STU. An example of a sample item together with transliterations of each alternative and their closest area of meaning is given here.

> The first man to produce a practical steam engine was Thomas Savery, an English engineer (1650-1715), who obtained a patent in 1698 (for a machine designed to drain water from mines). The machine contained no moving parts except hand-operated steam valves and automatic check valves, and in principle it worked as follows: Steam was generated in a spherical boiler and then admitted to a separate vessel where it expelled much of the air. The steam valve was then closed and cold water allowed to flow over the vessel, causing the steam to condense and thus creating a partial vacuum.
>
> 1. *Bokhar mishod tolid dar yek makhzane bokhar va rah yaft be yek luleye joda jaee ke an kharej kard bishtare hava.*[Steam is generated in a steam tank and then entered into a separate vessel where it expelled much of the air.] **Word for Word**
>
> 2. *Bokhar tolid mishod dar yek jush konandeye koravi ke be yek zarfe joda konande vasl shode bud va meghdare ziyadi hava as an kharej mishod.*[Steam is generated in a spherical boiling device which was attached to a separate vessel and a considerable amount of air was coming out.] **Reading Comprehension**
>
> 3. *Bokhar dar digi koravi tahiyye mishod va angah be zarfe digari hedayat mishod ke meghdare motanabehi hava ra ba feshar aghab mirand.*[Steam was generated in a spherical boiler and then admitted to a separate vessel where it expelled much of the air.] **Correct**
>
> 4. *Bokhar dar digi koravi ke be zarfe digari vasl mishod tahiyye shod ke meghdare motanabehi hava ra ba zoor birun kard.*[Steam in a spherical boiler attached to another vessel was generated that pulled out a considerable amount of air by force.] **Deviant**

2.4.2.4.1 Item analysis

To discard and/or revise items that were either too difficult or too easy, the researcher used the classic item analysis technique with the typical range of 0.33 to 0.67. Of the original 50 test items only 20 items remained to fit the standard item analysis range.

2.4.2.5 Post-test with sample population of students

After the necessary revision and clarification of the items, the final version of the translation test was prepared to be administered together with the Michigan test to another group of Electronics students. The testees were 60 male and female students from STU who were randomly selected from among 150 Engineering students.

3. Results

Based on the research questions stated earlier in this paper, statistical analyses were performed. The results for reliability, validity and factor analysis are given below.

3.1 Reliability

Reliability is defined as the extent to which a test produces consistent results under similar conditions with similar subjects. There are various statistical methods for measuring the reliability coefficient of a test (see Hatch and Farhady 1982). One of the most commonly-used ways of determining the reliability coefficient is the measure of internal consistency. In this study, in order to determine the reliability of the translation test and the subtests of the criterion measure, the measure of internal consistency (Kuder-Richardson formula 21) was used. As can be seen in the table below, the reliability of the translation test is lower than that of the subtests of the criterion measure. One of the most important factors which influence the reliability of a test is the number of test items: the more items used in a test, the higher the reliability of that test will be. Taking this into consideration, the main reason for the somewhat lower reliability coefficient of 0.74 may be the insufficient number of test items (the final version of the translation test consisted of 20 items which in comparison to the total 100 test items of the criterion measure is rather few). This being so, the translation test would probably have had a higher reliability coefficient if more items had been used. However, even the reliability coefficient actually achieved is satisfactory and encouraging.

Table 1. Reliability coefficients of the study measures

Subtests	
Grammar	0.90
Vocabulary	0.92
Reading Comprehension	0.93
Translation	0.74

3.2 Validity

Validity is defined as the extent to which a test measures what it is claimed to measure. To determine the validity of the translation test, correlational analysis was carried out. The concurrent validity of the translation test, as can be seen in Table 2., was low and not significant. In attempting to account for this, it should be pointed out that the coefficient of validity is influenced by many factors, including the size of sample. The greater the number of subjects taking a test, the higher the correlation coefficient of test results will be. This being so it is likely that one of the main reasons for the apparent low correlation of the translation test with the subtests of the criterion measure is the restricted sample of students who took the test (N=60). The correlation coefficient of the two tests might have been increased if a larger sample of test-takers had taken the test. It is also worth mentioning that the translation test and the criterion measure are fundamentally different from each other in terms of the purposes for which they are designed. Whereas the EFL criterion Michigan Test is primarily designed to assess the general language proficiency of the testees irrespective of their field of study, the newly developed translation test is mainly constructed for a specific group of students, namely students of Engineering and more specifically students of Electronics.

While both the criterion measure and the translation test are measures of language proficiency, the latter is more specific in that it claims to assess the language proficiency of the EST university learners. Therefore, it could be argued that there is something specific to the translation test which is not shared by the subtests of the criterion measure and that is the specific variance of the translation test.

Table 2. Correlation coefficients between the translation test and other subtests of the criterion measure

Variable	1	2	3	4
Grammar	*			
Vocabulary	0.27	*		
Reading Comprehension	0.24	0.30	*	
Translation	0.44	0.29	0.20	*

3.3 Factor analysis

Factor analysis, as Hatch and Farhady (op. cit.) point out, is based on the assumption that in any test there are probably one or more underlying traits being assessed. Through factor analysis the information on factors underlying a test is obtained by examining the common variance among items. Using the varimax rotation procedure in the SPSS computer package, the following data were obtained.

Table 3. Varimax factor matrix

Variable	Factor 1	Factor 2
Translation	0.54294	0.49639
Grammar	0.64303	0.48268
Vocabulary	0.83213	-0.16086
Reading Comprehension	-0.04363	0.86164

The data show us that there are loadings on factor 1 with vocabulary, grammar and translation. Factor 2 is heavily loaded with reading comprehension and moderately loaded with translation and grammar. Factor 2 and factor 1 contribute negatively as underlying factors for the vocabulary and reading comprehension respectively. The most crucial step in the interpretation of the above matrix is that of labelling these factors. It can be observed that factor 1 is highly loaded with grammar and vocabulary while reading comprehension contributes negatively to factor 1. Due to the function of the grammar and vocabulary tests which are considered to be discrete items, factor 1 could be labelled the discrete factor or comprehension of smaller chunks of language. On the other hand, factor 2 contributes negatively as an underlying factor for the vocabulary and is heavily loaded with reading comprehension and to some degree with grammar and translation. Given the integrative purposes for which reading comprehension passages are devised, and the negative load of vocabulary as a discrete item on factor 2, the second factor may be labelled integrative factor or comprehension of larger chunks of language. Factor 2 is also loaded with grammar, a discrete item type. This is probably due to the fact that grammatical knowledge is required for understanding a piece of text, namely, reading comprehension passages.

Taking the translation variable into account, it appears that factor 1 and factor 2 both contribute, if not highly, at least moderately to the translation. Thus, on this interpretation of the factor matrix the translation test may be labelled both as a discrete item and an integrative one.

4. Conclusion

The potential contribution of neglected translation methodology to ELT has recently been re-assessed. While translation methodology has been influenced by improvements in translation theory, its testing counterpart has been less enriched. The main purpose of this project was to develop procedures for the construction of an objective translation test. The procedures were designed to eliminate the possibility of subjectivity in the test and to achieve one of the essential properties of an objective test, called scorability. Compared with some batteries of language testing methods (mainly discrete tests (DP) and integrative tests (IN)) the translation test developed in this study has some advantages. Firstly, the translation test does not have the deficiency of the DP test, which has been criticized for not being able to take into account extra-linguistic factors (see Oller 1976); rather it is constructed at the level of a meaningful coherent unit of discourse. This means that every example of a rhetorical function used in this study has the property of being used in a natural context. Therefore, the translation test developed in this study does not violate the assumption of 'incoherent segments', the outstanding negative property of DP tests. Secondly, the translation test does not have the problem of independence of items which has raised doubts about the reliability of the cloze test (see Farhady 1980). Thirdly, through factor analysis, it has been shown that the translation test devised in this study can function not only as a discrete point test but also as an integrative test. Accordingly, the translation test can be supposed to assess both skills relating to the comprehension of smaller chunks of language (i.e. grammar and vocabulary) and those which relate to the comprehension of larger chunks of language (i.e. reading comprehension).

Further investigations are needed to shed more light on translation testing methodology. However, in our attempt to objectify translation tests we should be careful not to underestimate the potential value of the so-called subjective tests. We must always remember that the real merit of a translation test lies in its authentic practice of rendering a text. By carefully designing an open-ended translation test and training translation raters as well as specifying various weighting or scores for different types of translation errors, we may achieve objectivity in translation testing methodology.

Acknowledgement

I wish to acknowledge my sincere thanks to Alan Davies for editing the first draft of this paper and particularly Cathy Benson for her insightful comments on this project. I alone am responsible for any mistakes in this paper.

References

Farhady H. 1980. Justification, development and validation of functional language testing. Unpublished PhD. dissertation. University of California.

Harris D.P. 1965. *Testing English as a Second Language*. New York:McGraw-Hill Book Company.

Hatch E. and H. Farhady. 1982. *Research Design and Statistics for Applied Linguistics*. Rowley, Mass: Newbury House Publishers Inc.

Lado R. 1961. *Language Testing*. New York: McGraw-Hill Book Company.

Newmark P. 1981. *Approaches to Translation*. Oxford: Pergamon.

Nida E.A. and C. Taber. 1982. *The Theory and Practice of Translation*. Leiden: Brill.

Oller W.J. 1976. *Language Tests at School*. London: Longman.

Titford C. 1983. Translation for advanced learners. *ELT Journal*. 37/1: 184-90.

Trimble L. 1985. *English for Science and Technology : A Discourse Approach*. Cambridge: CUP.

Tudor I. 1987. Using translation in ESP. *ELT Journal*. 41/4.

THE SIGNIFICANCE OF "SIGNIFICANCE"

Martin Gill (DAL)

Abstract

Testing for statistical significance is an integral part of the methodology of research in applied linguistics, yet its implications are easily neglected. This paper examines some of them. After considering methodological issues raised by two examples from the literature, it proceeds to look in detail at a variety of misunderstandings attached to the reporting of "significant" results. Its conclusion is that significance testing is, at best, of limited utility, but, as commonly used, highly misleading. A final section considers the implications of abandoning the significance test.

1. Scientific versus statistical reasoning: two case studies

1.1 Introduction

Applied linguists, like other researchers in the human sciences, typically look to experiments to provide the hard data necessary to corroborate their hypotheses. Despite regular calls for greater use of qualitative procedures, drawn chiefly from ethnography, the field's dominant methodological paradigm, or the one to which it aspires, continues to be that of experimental science. As in physics, results obtained by experiment, properly insulated against sources of error, are taken (ideally) to permit valid inference to principles operating in the world; and the designers of such experiments rarely hesitate to claim that a decision to reject the null hypothesis (H_0), triggered by a result at the .05 level of significance, substantively strengthens not only the particular alternative hypothesis (H_1) they favour, but the theory from which it was derived. They may also claim to have inaugurated a promising research programme, to be carried forward as a matter of urgency with further detailed study, replication with larger samples, etc.

It is the aim of this paper to examine the soundness of such convictions; in particular to note points at which an analogy with physical science may misrepresent the experimental activity of applied linguistics, and so introduce the potential for distortion into the design and interpretation of applied linguistics research. This will involve taking a close look at the nature of the test for statistical significance and various false inferences that may be drawn from it. The paper also attempts to show how a more general tendency to regard statistical procedures as scientific instruments for uncovering independently existing empirical phenomena obscures the more basic question of what research objects are, and the role of a methodology in defining them. It is argued that, to the extent that a methodology is constitutive of the kinds of knowledge it makes possible, the notion of 'independently existing phenomena' cannot be sustained. We turn first to two examples which serve to illustrate some of these issues.

1.2 Supporting a theory (Hafiz and Tudor 1989)[1]

Hafiz and Tudor report the successful results of a 12-week reading programme 'inspired by Krashen's Input Hypothesis' (Hafiz and Tudor 1989:4) which they conducted using simplified readers with a class of young Pakistani ESL learners: on post-testing, the experimental group achieved statistically significant improvements on their pre-test scores on all parts of a battery of reading and writing tests, while, for the most part, the two control groups did not. This outcome may not strike the outsider as especially surprising, given that the participants in the programme had devoted on average 42 extra hours to English, not counting the reading they did at home (ibid.:7), whereas the control groups had merely followed their usual classes. Yet for Hafiz and Tudor it

> lend[s] support to Krashen's Input Hypothesis, indicating that extensive L2 input in a tension-free environment can contribute significantly to the enhancement of learners' language skills, both receptive and productive.
>
> (Hafiz and Tudor, op. cit.:10)

Remarks of this kind are common enough in the literature. They create an impression of scientific progress, of theoretical understanding fortified by the accumulation of well-attested empirical results. But they ignore crucial differences between the logic of scientific reasoning and that of the probabilistic reasoning (as institutionalized in the t-test and its more elaborate extensions) from which the data they are concerned with are derived. These differences, therefore, deserve to be clearly stated.

What is wrong with the conclusion Hafiz and Tudor draw from their experiment has nothing to do with the plausibility of the suggestion that tension-free reading can help learners to become more proficient; nor is it just a consequence of the shakiness of their experimental design, although of course that is relevant to judging the validity of their research as a whole (cf. section 3.4 below). It is the invalid inference they make (or imply by their choice of words) from specific observation to general theory. It is a mistake that is easily overlooked in a procedure which applies statistical significance tests for the purpose of adjudicating between hypotheses, for this practice obscures the fact that the decision so determined is logically independent of any inference to the truth or otherwise of the theory from which the hypotheses are deduced, or of any attempt to attach a degree of confirmation or even remote probability to it. The results Hafiz and Tudor obtained may well be consistent with Krashen's Input Hypothesis, but they lend no more support to it than to any other plausible (or, for that matter, any wholly absurd) theory with which they might also happen to be consistent.

It is easy to see that this experimental programme was designed to illustrate a conclusion to which the researchers were already committed. Probably this increased the practical value of the programme for the group that took part in it, but it negates any scientific claim to have tested a theory, or corroborated general principles of language learning. If this escaped the researchers, it was perhaps because the significance test procedure in itself appeared to be sufficient guarantee of a scientific outcome. Notwithstanding the absence of the 'testable and falsifiable universal laws and initial conditions' that Popper (1979:193) sets as a precondition of scientific explanation, an experimental result achieving statistical significance is likely to seem (especially to anyone already persuaded) a persuasive indication of the existence of a fact of "genuine", intrinsic significance (we may speculate which sense of this word is intended in the quotation above). This can occur, as here, regardless of doubts about the experimental design and sampling procedure, regardless of the looseness of the operational definitions chosen, and regardless of how short of specific predictive power the theory in question may be.

It would seem that any analogy between this procedure and physics must be mistaken; the credibility of physical theories is not increased by decisions of the sort determined by significance testing. But nor is it increased by observing a non-zero difference in the predicted direction between experimental and control conditions, at least not if the observation in question is entirely consistent with everyday expectations (for example, that learners learn better when they are relaxed, interested, etc.). What is required of a theory is a capacity to make novel predictions which can be subjected to exact empirical scrutiny. A theory can be said to be strengthened, at least our belief in it can be said to be more adequately justified, the longer it survives the closest scrutiny we can give it. This presupposes a theory with some interesting empirical content, capable of refutation: to the extent that Krashen's is not such a theory (as Gregg (1984) and McLaughlin (1987) argue it is not), Hafiz and Tudor could not have hoped to lend it a crumb of support, whatever their method.

Given the inherently inexact nature of their subject, very few theories in the behavioural sciences are likely to measure up to these "scientific" criteria, for example by successfully predicting the size (not just the direction) of a difference. The main justification for using the significance test is to fill this absence of precision by supplying a way of deciding when an observed value is unlikely to have occurred by chance. However, the dangers exemplified here are, first, that "achieving significance" may seem to play a role in the logic of theory-testing and the rhetoric of research-paper writing that is equivalent to that of overcoming the much more demanding observational hurdles usual in physical science; and, second, that a theory without substance will be dignified, and its position consolidated, by the published announcement of "confirmation". The logical problem of confirmation is discussed further in section 3.1, and its relation to theory-testing in the physical sciences is developed in 3.5.

1.3 Exploring a concept (Ferguson and Maclean 1991)[2]

Experiments in the behavioural and human sciences, including applied linguistics, often proceed in an "exploratory" fashion, without a theoretically motivated design, but trusting to statistical techniques to reveal what phenomena are of interest. Given the complexity of many of these techniques, it becomes tempting to view this as purely an instrumental matter, the atheoretical application of sophisticated tools to get at the underlying constituents of reality (the "facts") on the basis of which a subsequent theory will be constructed. Here too there may persist some imagined parallel with what physicists do. Not only is this image misleading, however (for physical no less than for statistical sciences), it also leaves the experimenter unguided as to which phenomena might be genuine, and which simply artefacts of the chosen method. As John Dewey observed:

> A quantitative statement with no theory to determine what is being measured would justify calling the "measuring" of all cracks in the plaster of my wall "science" if it were done with elaborate statistical technique.
>
> (Dewey 1949; cited in Johanningmeier 1980:54)

In a recent study, Ferguson and Maclean (1991) seek to analyse the properties of subjective judgements of (medical) text difficulty, and, in particular, by Principal Components Analysis, 'to get below the surface of things' (op. cit.:123) to establish the (true) dimensionality underlying the seven explicitly formulated categories of difficulty used by their team of judges in assessing texts. It emerges from the computation that there are just 'two significant dimensions' of difficulty (ibid.:122), which therefore, it seems, are to be regarded as the real, unconscious causes of the judges' conscious behaviour: 'Perhaps, then, the judges were in fact operating with two dimensions though they may have believed they were independently assessing seven' (ibid.). Applying the statistical procedure has not merely revealed broad patterns of co-occurrence in the data, but got at hidden facts which are in some sense intrinsically more explanatory than those on the surface (this may explain why the writers do not report the views of the judges themselves about the judging task). Moreover, the analysis assumes that the sense in which these facts are more explanatory is cognitive, equating their hiddenness in the data with the hiddenness of mental activity in the heads of the judges.

We might recall J.S. Mill's warning about the dangers of reification:

> The tendency has always been strong to believe that whatever received a name must be an entity or being, having an independent existence of its own. And if no real entity answering to the name could be found, men did not for that reason suppose that none existed, but imagined that it was something peculiarly abstruse and mysterious.
>
> (J.S. Mill cited in Gould 1981:320)

What is striking in this context is that the entities in question had no name before this particular study found them, but that, even so, their "reality" was assured in advance by their emergence from statistical analysis. The scientific challenge lay in establishing their correct identities: 'technically speaking, they stand in need of reification' (Ferguson and Maclean, op. cit.:122). Accordingly, the authors conjecture that the first principal component represents 'general language difficulty', but remain doubtful about the second ('something to do with contextual support and rhetorical organization' (ibid.)).

Here again, it is important to be clear what the issue is. It is not in question that for the practical purposes (i.e. efficiency of text grading) that are the experimenters' immediate object (ibid.:118), discarding the unreliable and hardly quantifiable variables that contributed least to the overall assessment of difficulty is obviously sensible. Nor are the statistical procedures necessarily suspect in themselves. The problem appears when these procedures are used in the analysis of the underlying causes of the difficulty judgements, simultaneously to provide a conceptual model of the judgements (they "really" consist of two components), and empirical evidence to support it (their occurrence in this study), for in this way the method tends simply to confirm the validity of its own artefacts. Moreover, instead of achieving clarity of understanding, we are left facing a paradox. the statistical method is taken to have delivered a deeper,

71

more real picture of the relevant cognitive activity than the consciously evolved, subtle descriptions of the agents themselves, yet it proves, on inspection, to be devoid of interesting content. To the writers, the natural solution is to indicate the need to refine and elaborate their statistical analyses (ibid.:123). Without a theoretical model that will permit interpretation of the results independently of the method used to derive them, however, these refinements will be of little use.

1.4 Conclusion

The studies examined here illustrate in different ways the readiness of experimenters to use inferential statistical methods, independently of a theoretically conceived research design, to do the work of conceptual analysis, by-passing experienced judgement, and licensing the affirmation of general theoretical conclusions. The remainder of this paper considers these misconceptions in greater detail. The following section first outlines doubts raised about the place of significance testing in research in the behavioural and social sciences.

2. Questioning the significance test

2.1 The emergence of doubts

Insisting that results achieve a specified level statistical significance (e.g .01) has sometimes been used by editors as a means of preventing the literature from overflowing with spurious studies (see, for example, Melton quoted in Bakan 1966:426f). Yet by itself this does not hold back the tide. If anything, it inclines experimenters to publish claims for hunches apparently "confirmed", but to discount null hypotheses left unrejected; while if subsequent evidence then points to the truth of H_0, this will tend to go unreported (Carver 1978:396). The result, as noted above, is that significant progress is publicly announced where the physical sciences might at best see only 'private clues for future exploration' (Hogben 1970:19). More generally, attaching undue emphasis to statistical significance encourages experiment where none is justified, just because the test is easy to apply and creates an impression of scientific objectivity. So, for example, Venezky's survey of research into reading instruction published in the United States notes that the bulk of it is composed of meaningless statistical exercises, 'fishing expeditions ... almost random searches for relationships, unanchored by any theoretical frameworks and often unbothered by the limitations of the methods employed' (Venezky 1984:17); 'an enduring testimony to the patience of the American printer and the vulnerability of American forests' (ibid.).

Similar misgivings surfaced during the 1960's in the American psychological research literature, prompting a re-examination of the role played in it by statistical significance testing, and leading to conclusions such as Lykken's that

> statistical significance is perhaps the least important attribute of a good experiment; it is *never* a sufficient condition for concluding that a theory has been corroborated, that a useful empirical fact has been established with reasonable confidence - or that an experimental report ought to be published.
>
> (Lykken 1968:158; original emphasis)

The particular object of criticism was professional ignorance of the logic and interpretation of the tests, and the widespread readiness to make invalid inferences based on them (Rozeboom 1960; Bakan 1966; Meehl 1967; Lykken 1968), a readiness for which empirical studies have provided evidence (see, for example, Rosenthal and Gaito 1963; Kahneman et al. 1982). The discussion extended to philosophical misgivings about the status of statistical inference in a scientific paradigm, reflecting disputes, described by Hogben (op. cit.), of longer standing within statistical theory itself about the true purpose and scope of the test. The psychological papers cited above, together with others from sociology and elsewhere, have been collected by Morrison and Henkel, who summarize the general consensus:

> The significance test as typically employed in behavioural science is bad statistical inference, and ... even good statistical inference in basic research is typically only a convenient way of sidestepping rather than solving the problem of scientific inference.
>
> (Morrison and Henkel 1970:xi)

The same conviction regarding educational research has since been forcefully expressed by Carver:

> The emphasis on statistical significance over scientific significance in educational research represents a corrupt form of the scientific method. Educational research would be better off if it stopped testing its results for statistical significance.
>
> (Carver 1978:378)

There is little sign, it must be said, that research practices in these fields have responded to such criticism; obstacles to adopting the solution Carver proposes are discussed in section 4.2.

2.2 The case of applied linguistics

Examining the issues raised by these critics as they relate to research in applied linguistics will highlight aspects of the relationship between probabilistic techniques and other kinds of inductive reasoning that are easily glossed over, and dispel any expectation that an empirical approach will in itself lead to clearer understanding, or that 'if our treatment of our subject matter is mathematical it is therefore precise and valid' (Bakan op. cit: 437). Far from being of exclusively philosophical interest, this activity should be seen, in Hogben's words, as 'the birthright and duty of every scientific worker who subjects his data to ... statistical inference' (Hogben, op. cit.:14-15).

Nevertheless, few applied linguists, at least among those from "arts" backgrounds, will feel qualified to evaluate the statistical techniques taught as the routine practice of the discipline, still less their adequacy within the experimental paradigm. It is easier to trust expert assurances that they "work", as one can learn to drive without understanding about cars. Since, moreover, the results turned out by these techniques will take the form preferred by the research community, it may never seem urgent to criticize their presuppositions. By these means, researchers are socialized into treating the statistical significance test as a paradigm of scientific rationality and ready-made inferential device sufficient to all normal purposes. Indeed, doubts of the kind expressed in

1,000 we might be wrong ...' (Hatch and Farhady 1982:106); by 'wrong' they presumably mean 'committing a Type I error' (see below); but it is hard to resist the implication that there are 999 chances in 1,000 that we (and our theory) are right.

Furthermore, it will be observed that rejecting the coin result as valid evidence for T on the grounds that it obviously must have arisen by chance (etc.) is equivalent to accepting the result of the reading programme above as evidence for the Input Hypothesis, on the grounds that the idea seems plausible. Both cases involve stepping outside the test procedure to draw on reasons derived from other sources, such as our knowledge of the world, our expert judgement, or our commitment to a theoretical position: to the extent that such sources are available and decisive (as in science they should be), we do not need statistical significance tests as "decision mechanisms" at all.

3.2 Chance

p expresses the probability of our committing a Type I error: i.e. of falsely rejecting H_0. There is nevertheless a natural inclination among experimenters, reinforced, as Carver shows, by the writers of introductory statistical texts, to regard p as a statement of the 'odds against chance' (cf. Carver, op. cit.: 383). In its eagerness to get from statistical significance to corroborating conclusion, this is cognate with the belief in automatic inference. If p is held to express how likely it is that the result may have turned up by chance, its reduction to negligible levels implies that something substantial has been caught (presumably evidence for T) in the experimental net. In reality, of course, the significance test is premised on the truth of H_0, i.e. that chance ("sampling error") alone accounts for the observed result (and, it is worth noting, this assumption must be adhered to in practice by sampling at random from the population in question; if this is not done, for example where "convenience" samples are used, like the local school classes in the Hafiz and Tudor study, statistical significance can have no serious meaning (cf. Hewitt 1982:16)).

However, it is one thing to accept that results which are improbable under H_0 will occasionally turn up, so that over time a small percentage of published experiments will contain Type I errors, i.e. wrongly accept H_1 (ideally 5%, with $p=.05$ as the acceptance criterion, although pressure to select significant and ignore non-significant results will tend to push the number up). It is another, and something we cannot usually hope to determine, to say what the odds are in any given instance that a Type I error has occurred. In other words, the logic of significance testing allows us only to talk about central tendencies in the population of experiments, not about single instances. For this reason alone (independent of logical considerations) it would always be wrong to interpret a single result in the acceptance region as support, etc. for a hypothesis. Moreover, without considering the power of experiments, it is impossible to guess the likelihood of their having uncovered a "genuine" phenomenon, however minute the level of significance achieved (see 3.4 below).

3.3 Replication

One hallmark of a strong scientific theory is its resilience under repeated experiment. It must hold universally, subject to the calculable influence of other variables (plus various simplifying assumptions), not just for some favoured group of experimenters, or in some privileged location. If applied linguistics claims to deal with universal principles of language learning in this sense, its theories, too, should withstand replication. In fact, the necessity is the more urgent, in one sense, given the relatively high proportion of spurious results the significance test procedure will instruct us to accept (presumably the spirit theory for coins will not survive a second trial).

In practice, replication tends to be neglected. It is true there is a genuine difficulty, faced by any behavioural discipline, of knowing exactly how replication is to be understood; for although physical laws may be independent of time, place, culture, etc., human behaviour clearly is not, so that it will be unreasonable to hope for more than a rough identity of conditions between the different occasions on which the "same" experiment is performed. But, as with refutation, the main reason appears to be that significance testing assigns no role to it; in fact, the test rules out replication a priori. As Bakan points out, the 'once-ness' of an experiment is a condition of the inferential model on which the test as Fisher conceived it is based; its logic presupposes an infinite hypothetical universe, of which the actual experiment represents a random sample, and it will be undermined by replication unless the probabilities are adjusted so as to treat both as a single entity (Bakan, op. cit. 424-5; cf. Hagood 1970:67). Thus, a succession of experiments designed to test the same theory, each achieving statistically significant results, cannot be regarded as a substitute for replication, whatever the temptation to do so. Conversely, we may take the view that a given set of behavioural data cannot be separated into its universal essence, and the effect of other variables and simplifying assumptions just referred to, in other words that it is uniquely shaped by its context and therefore non-replicable. In this case, if we test for significance, we shall need a clear understanding of just what 'infinite hypothetical universe' is intended (Hagood op. cit.:70).

3.4 Power

The difficulty of knowing what, if anything, is a substantive phenomenon in our experiments, and of being sure that it is such a phenomenon that an experiment has uncovered, raises the further difficult question of experimental power. In the natural sciences power will be a matter of instrumentation: finer scales, better lenses, etc. In the human sciences, including applied linguistics, it will depend principally on strength of experimental design and on sample size.

For practical purposes, if we want to discover the existence of real entities, rather than the non-existence of unreal ones, experimental power should interest us. Yet the significance procedure disregards it, except insofar as the prior determination of a critical level of significance is a trade-off between the acceptability of Type I errors and those of Type II (i.e. failure to reject a false H_0). While we can try to decide for ourselves whether an experimental design is valid, the fact remains that emphasis in published results on exceptionally small values of p diverts attention away from weaknesses. Carver calls this the 'replicability or reliability fantasy' (Carver: op. cit.:385). As long as (1-p) is taken, consciously or otherwise, to express the reliability of the result obtained, it will appear, quite unjustifiably, to validate *post hoc* whatever design has been used: crudely, that if it has a "highly significant" label attached to it, it must have been a good experiment. This also helps to reinforce the idea that statistical significance can stand in lieu of replication, with progressive research in a given field seen as the accumulation of results so labelled (cf. above).

Why this will not do will be discussed further in a moment. But the reason for it no doubt reflects the relative ease of calculating statistical significance, as against the relative difficulty of calculating experimental power. In applied linguistics experiments, "noise" levels are generally high, so that there is no sure way of knowing if the effect observed is "really" the effect being looked for. Unlike typical scientific variables, which may be predicted, isolated and measured with great accuracy, the variables which interest us (L2 proficiency, reading comprehension, difficulty judgements, etc.) depend critically on the validity of a series of secondary inferences. They may be spoken of as independent entities - like 'sheer comprehension', for example, which the experimenter hopes to distinguish from memory, inference, deduction, reasoning, intelligence, etc. (Carroll 1972:3ff) - but their identities have themselves emerged from statistical procedures with quantities (such as test scores) whose interpretation is open to debate, and interact non-randomly with a host of variables in the educational, psychological and cultural backgrounds of the subjects. Instead of making point predictions, our hypotheses deal with directional tendencies in population means, inferable only on the basis of (frequently small) samples; there is no ready way of predicting how big observed effects should be, and no independently interpretable scale on which to represent them.

It would be wrong to expect obtained levels of significance to fill all, or any, of these requirements, given the ease with which they can be manipulated by the experimenter, especially where sample size is concerned (see below). On the other hand, for normal, "messy" experimental situations power cannot be calculated in advance. Therefore, we cannot know how often we fail to detect a difference that really exists, in other words commit a Type II error. The proportion of Type II errors in the population of applied linguistics experiments must nevertheless be presumed to be rather high, since the probability of Type II error - which we may call p_2 - will be inversely proportional to the probability of Type I error, p_1 (or, simply, p): i.e. the fewer Type I errors we allow (by being reluctant to reject H_0), the more Type II errors we will make (by letting through H_0s that are in fact false), and vice versa. Our aim will be to tighten experimental design towards the ideal point of no noise, where $p_2=0$, but, as we have seen, noise is an irreducible property of the variables we want to investigate. Power, expressed by $(1-p_2)$, must therefore frequently be low.

According to Tversky and Kahneman (1982) in a study of how probabilistic data are interpreted by those who use them, the worst of the 'self-defeating' and 'pernicious' consequences of low experimental power is not so much the valid but discarded hypotheses strewn along the pathway of research, as the readiness on the part of experimenters, for which they quote evidence from a survey of research psychologists, to explain noise (ibid.:27), to seek causal explanations for unexpected differences between the results of an experiment and its attempted replication, when it is quite beyond the power of the experiment to resolve them into true and chance effects. In this way, experiments constantly add spurious facts to the repository of knowledge, and equally spurious theories may be evolved to explain them.

3.5 Power and "corroboration": a paradox

A paradox discussed by Meehl (1967) concerning the nature of experimental power illustrates the true distance between hypothesis testing in behavioural and natural sciences. It rests on the reasonable assumptions that (1) the aim of any experimenter will be to improve experimental power towards 100%; and that (2) H_0 is almost always false in any population; in other words that any treatment (e.g. extensive reading) or criterion of classification (e.g. sex, father's religion, etc.) will have some influence on output measures that will be detectable given sufficient power, or, which is the same thing, a large enough sample (see also 3.6). Meehl's argument adopts, as a limiting case, the further assumption (3) that all our theories are equally unlikely. If assumption (2) is true, it follows, quite independently of other considerations, that, over the infinite set of possible experimental or quasi-experimental situations in (say) applied linguistics, there will be a non-zero difference between "experimental" and "control" groups on the variable of interest in practically every case (and for the purpose of this argument it does not matter which group is so designated). Assuming that these differences are normally distributed, the difference observed will be in the direction that favours the experimental group in 50% of cases. Let this outcome (again arbitrarily) be called "success", and that in the other 50% of cases, in which the difference favours the control group, be called "failure". If all these experimental situations are then paired off randomly with theories drawn from the infinite set of real or potential theories in applied linguistics, there is a 50% probability that a theory, however wrong "in the state of nature", will find itself paired with a "successful" experiment. In other words, the consequence of our experimental method is to yield a prior probability equal to 50% of finding experimental "support" for any of our theories (ibid.:113).

Meehl is at pains to stress that this is a limiting case, 'a lower bound on the success-frequency of experimental "tests"'(ibid.:111), given assumption (3) above, and assuming "perfect power" (i.e. certain detection of any difference that exists). He therefore concludes that, paradoxical as it may seem, any attempt to increase the power of experiments in the real world will only make the "observational hurdle" for a theory easier to overcome, by bringing the probability of achieving "corroboration" ever closer to 50%, even where the theory in question is intrinsically worthless. In the natural sciences, by contrast, increasing experimental power achieves just the reverse: as calibrations and measurements gain in precision, so theories are forced to pass progressively more stringent tests, reducing towards zero the chances of survival for any but the very fittest.

When combined with the ever-present temptation, discussed earlier, to confuse rejection of H_0 with confirmation of T, Meehl's paradox shows how easily even an apparently fruitful research programme might come to be based entirely on a self-perpetuating chain of flawed statistical inferences.

3.6 The fiction of the null hypothesis

The only conclusion that can legitimately be drawn from a statistically significant result is that there is a probability equal to the obtained value of p that H_0 was wrongly rejected. As Hogben comments:

> For what reason ... should [the researcher] be eager to take advantage of a test which can merely assign a low probability to erroneously asserting that the treatment is useless ... ? ... The terminal statement which the test procedure ostensibly endorses provides an answer (if any) devoid of operational value in the context of an experiment rightly undertaken to confirm a positive assertion suggested by prior information. Since the test procedure merely endorses the negation of a null hypothesis conceived within the straitjacket of the single infinite hypothetical population, the outcome will thus be an irrelevant decision or no decision at all.
>
> (Hogben, op. cit.:35)

It has been the object of the argument to this point to illustrate the hankerings that are widely felt for some surer mechanism to generate true substantive statements; hankerings which are readily but illicitly projected on to the significance test. No doubt these dangers are well-known: it would be an elementary mistake to attach to p any interpretation other than the one given above. It is unlikely that researchers will cling to just one of the false notions discussed; but they may be inclined, if only under the influence of common usage, at different times to fall into any of them. The most serious problem arises when the level of significance expressed by p is made to bear the weight of a decision to regard a piece of research as deserving further attention.

The thrust of Hogben's criticism here, however, is aimed less at the feebleness of assertion made possible by the significance test than at the fiction of the null hypothesis (Carver's 'straw man' (op. cit.:381)) upon which it is premised. Not only can we never claim to have "confirmed" H_0 on logical grounds, but to seek such confirmation would be irrational a priori. Researchers do not set up experiments in the belief that their treatment has no effect whatever, that the mean difference between an infinite number of sample scores drawn at random from an infinite population, treated and untreated, will be precisely 0.00. Given the complexity of the variables of interest to us, the probability of such a result is vanishingly small. It is safe to say that H_0 is never true. Indeed, for cases in which H_0 has not been rejected, it is common simply to repeat the experiment with a larger sample. Bakan refers to his own tests on data collected from 60,000 subjects: if N is large enough, almost any difference (e.g. east *vs* west of the Mississippi, north *vs* south, etc.) can be shown to be "significant" (Bakan, op. cit. 425; cf. also Meehl, op. cit.:109). Therefore the notion of random sampling under H_0, which is essential to the calculation of p, corresponds to no conceivable state of the experimental situation, either on a single occasion, or (much less) over time (cf. Hagood, cited in Hogben, op. cit.:40). In these circumstances it would be hard to make a convincing case on scientific grounds for persisting in its use.

4. Conclusion

4.1 Decision versus interpretation

These remarks may be enough to establish that significance testing is not in any sense just a version of scientific procedure tailored for a behavioural discipline. Even when it is approached as a probabilistic mechanism of strictly limited utility, however, there remain unresolved anomalies in its use, traceable, as Hogben shows, to profound disagreement at the level of statistical theory. If it is regarded as a decision test, triggering acceptance of hypotheses that achieve a pre-set level of significance, then its function should only be to ensure stability in the incidence of Type I error across the aggregate of experimental results. As such, it will admit no interpretation of results considered singly, and no attempt to equate levels of significance with an experimenter's strength of conviction. Yet there is a widespread tendency, already noted, for significance tests to be applied for the purpose of establishing degrees of belief with respect to single results. Worse, the two approaches are regularly combined without thought for their divergent implications, making it appear that a series of positive test decisions actually entails increasing conviction, even though such factors as the critical level of significance and the size of sample chosen are arbitrary or matters of convention.

This is no way for science to proceed. As Rozeboom argues:

> A hypothesis is not something like a piece of pie offered for dessert, which can be accepted or rejected by a voluntary physical action. Acceptance or rejection is a cognitive process, a *degree* of believing or disbelieving which, if rational, is not a matter of choice but determined solely by how likely it is, given the evidence, that the hypothesis is true. ... While the scientist - i.e. the person - must indeed make decisions, his *science* is a systematized body of (probable) *knowledge*, not an accumulation of decisions.
>
> (Rozeboom, op. cit.:423; original emphasis)

Since experimenters have a duty to interpret their findings, and since their aim must be to establish a sound basis for the growth of knowledge in their field, the moral to be drawn is that significance tests, in any of their guises, are at best weak, at worst inappropriate and misleading. Lakatos, viewing these matters from the perspective of natural science, puts the point less charitably:

> One wonders whether the function of statistical techniques in the social sciences is not primarily to provide a machinery for producing phoney corroborations and thereby a semblance of "scientific progress" where, in fact, there is nothing but an increase in pseudo-intellectual garbage.
>
> (Lakatos 1978:88, n.4)

4.2 Further implications

What would abandoning the test for statistical significance, or at least relegating it to a minor role in the analysis of data, mean for research in applied linguistics? The chief obstacle to dispensing with the significance test is that without it the research enterprise would seem to founder; not because there are no alternatives, but because that enterprise is, to a great extent, premised on the test and the kind of the knowledge it makes available. As Bakan has put it:

> [The test of significance] is profoundly interwoven with other strands of the ... research enterprise in such a way that it constitutes a critical part of the cultural-scientific tapestry. To pull out the strand of the test of significance would seem to make the whole tapestry fall apart.
>
> (Bakan 1966:428)

The failure of qualitative methods to make headway against inferential statistics may be attributable in part to the fact that the former are likely to be viewed as anecdotal, short of "scientific" rigour and inadequately generalizable; in short, as not conforming to the notion of properly constituted knowledge in the field. For this reason, there exists no established research discourse in applied linguistics to which such methods can contribute. But at the same time, no doubt, this state of affairs has been maintained by the prestige of the physical paradigm, by the tendency of our culture 'to view the exact sciences as the long-sought description of the "true and ultimate furniture of the universe"' (Putnam 1981b:15). It is against this background that the present discussion, echoing the work cited in section 2.1, has sought to put in doubt the assumed scientific rigour of significance testing, by showing that in many cases it is illusory and its implications readily misinterpreted. It has argued that the use not only of the logic but also of the language of "corroboration" derived from physical science creates the impression of an empirical research programme progressing towards an ever clearer and better supported theoretical understanding, where no such impression may be justified.

It remains, however, that the special status of physical science naturally favours the belief that its representations are more nearly "true" in some absolute sense, so that it may still be taken for granted that better theoretical description of phenomena must mean, ultimately, closer approximation to the "objective" picture of the world delivered by physics. It is perhaps because the notion of "correspondence to the facts" is taken to be unproblematic that researchers even in the non-physical sciences have been able to develop highly sophisticated methodologies without giving equivalent attention to conceptual issues, treating methods as different kinds of tools, and choice among them as independent of the conceptualization of the empirical "facts" to be discovered. The further purpose of this argument has therefore been to suggest, on the contrary, that '"objects" do not exist independently of conceptual schemes. We cut up the world into objects when we introduce one or another scheme of description' (Putnam, op. cit.:52); that, as Hacking observes, 'a style of reasoning may determine the very nature of the knowledge that it produces' (Hacking 1981:143; cf. also his 1982:49ff).

For just this reason, it would be wrong to imply that probabilistic methods are intrinsically less valid than others. It is a matter of history that statistical modes of thought have increasingly been perceived as explanatory in the human sciences, and have made it possible to trace interesting relationships among phenomena. The very idea of "the human sciences" owes its possibility to advances in probabilistic techniques and the emergence of styles of reasoning associated with them in the nineteenth century (see, for example, Porter 1986, Stigler 1986, Hacking 1990). But to the extent that statistical methods are supposed, in the paradigm of physical science, to reveal (for example) the hidden facts of human cognitive operation, it is necessary to question the more or less automatic use that is made of them in our field.

Abandoning the significance test is not therefore just a methodological problem, or a matter of personal preference, as if we might replace it with something perhaps 'softer' and more congenial but in other respects continue with the work we are doing. If we accept that a methodology is (or necessarily implies) a style of reasoning, the change will essentially redefine that work itself, that is, the objects of research and the ways in which we think about them.

Acknowledgements

I should like to thank Robert Hill for his encouragement in the writing of an earlier version of this paper.

Notes

1. For a more circumspect analysis of the same results, see also Tudor and Hafiz (1989).

2. Perhaps it is unfair to discuss in detail research that is only reported in a working paper; however, attention here is directed less to its specific results than to the way the writers conceptualize one part of their project.

References

Bakan D. 1966. 'The test of significance in psychological research'. Psychological Bulletin 66/6: 423-37.

Carver R. 1978. 'The case against statistical significance testing'. Harvard Educational Review 48/3. 378-99.

Carroll J. 1972. 'Defining language comprehension: some speculations' in J. Carroll and R. Freedle (eds.) 1972. Language Comprehension and the Acquisition of Knowledge. New York: Wiley and Sons. 1-29.

Gould S. 1981. The Mismeasure of Man. New York: Norton.

Gregg K. 1984. 'Krashen's monitor and Occam's razor'. Applied Linguistics 5/2: 79-100.

Ferguson G. and J. Maclean. 1991. 'Assessing the readability of medical journal articles: an analysis of teacher judgements'. Edinburgh Working Papers in Applied Linguistics 2: 112-124.

Hacking I. 1981. 'Lakatos's philosophy of science' in I. Hacking (ed.) 1981. 128-143.

Hacking I. (ed.) 1981. Scientific Revolutions. Oxford: Oxford University Press.

Hacking I. 1982. 'Language, truth and reason' in M. Hollis and S. Lukes (eds.) 1982. Rationality and Relativism. Oxford: Blackwell. 48-66.

Hacking I. 1990. The Taming of Chance. Cambridge: Cambridge University Press.

Hafiz F. and I. Tudor 1989. 'Extensive reading and the development of language skills'. ELTJ 43/1: 4-11.

Hagood M. 1970. 'The notion of a hypothetical universe' in D. Morrison and R. Henkel (eds.) 1970. 65-78.

Hatch E. and H. Farhady 1982. Research Design and Statistics for Applied Linguistics. Cambridge, Mass.: Newbury House.

Hewitt G. 1982. 'A critique of research methods in the study of comprehension'. British Educational Research Journal 8/1: 9-21.

Hogben L. 1970. 'The contemporary crisis or the uncertainties of uncertain inference' and 'Statistical prudence and statistical inference' in D. Morrison and R. Henkel (eds.) 1970. 8-40. (Chapters reprinted from L. Hogben 1957. Statistical Theory. New York: Norton.)

Johanningmeier E. 1980. 'American educational research: applications and misapplications of psychology to education' in J. Smith and D. Hamilton (eds.) 1980. The Meritocratic Intellect: studies in the history of educational research. Aberdeen: Aberdeen University Press. 41-57.

Kahneman D., P. Slovic and A. Tversky (eds.) 1982. Judgement Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Lakatos I. 1978. 'Falsification and the methodology of scientific research programmes' in J. Worrall and G. Currie (eds.) 1978. The Methodology of Scientific Research Programmes: Philosophical Papers of Imre Lakatos 1. Cambridge: Cambridge University Press. 8-101.

Lykken D. 1968. 'Statistical significance in psychological research'. Psychological Bulletin 70/3: 151-59.

McLaughlin B. 1987. Theories of Second Language Learning. London: Arnold.

Meehl P. 1967. 'Theory-testing in psychology and physics: a methodological paradox'. Philosophy of Science 34: 103-15.

Morrison D. and R. Henkel (eds.) 1970. The Significance Test Controversy. London: Butterworths.

Popper K. 1959. The Logic of Scientific Discovery. London: Hutchinson.

Popper K. 1979. Objective Knowledge: an Evolutionary Approach (revised edition). Oxford: Clarendon Press.

Porter T. 1986. The Rise of Statistical Thinking 1820-1900. Princeton: Princeton University Press.

Putnam H. 1981a. 'The "corroboration" of theories'. in I. Hacking (ed.) 1981. 60-79.

Putnam H. 1981b. Reason, Truth and History. Cambridge: Cambridge University Press.

Rosenthal R. and J. Gaito. 1963. 'The interpretation of levels of significance by psychological researchers'. Journal of Psychology 55: 33-38.

Rozeboom W. 1960. 'The fallacy of the null-hypothesis significance test'. Psychological Bulletin 57/5: 416-28.

Stigler S. 1986. The History of Statistics: the Measurement of Uncertainty before 1900. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Tudor I. and F. Hafiz 1989. 'Extensive reading as a means of input to L2 learning'. Journal of Research in Reading 12/2: 164-178.

Tversky A. and D. Kahneman. 1982. 'Belief in the law of small numbers'. in D. Kahneman, P. Slovic and A. Tversky (eds.) 1982: 23-31.

Venezky R. 1984. 'The history of reading research' in P. Pearson (ed.) 1984. Handbook of Reading Research. New York: Longman. 3-38.

Wason P. 1968. 'On the failure to eliminate hypotheses. a second look' in P. Johnson-Laird and P. Wason 1977. Thinking: Readings in Cognitive Science. Cambridge: Cambridge University Press. 307-314.

Editor's note

Ferguson and Maclean were invited to respond to the comments on their work in this paper. Maclean does not feel it necessary to do so, Ferguson has indicated that he may respond later.

THE DESIGN AND VALIDATION OF A MULTI-LEVEL READING COMPREHENSION TEST

Aileen Irvine (IALS)

Abstract

Constructing a test of EFL reading comprehension which will accommodate the complete spectrum of performance from beginner to near native-speaker can be problematic. Such a test is currently being developed at the Institute for Applied Language Studies, University of Edinburgh. With only a few hundred students to validate such a widely discriminating test, the problems become practical as well as theoretical. This article is a short report on why there was a need for the test and how the problems were approached.

1. Background to the test

The Edinburgh Project on Extensive Reading (hereafter referred to as EPER) aims to offer a complete "extensive reading" package to its customers. That is to say that EPER not only organises the selection and dispatch to the client of appropriate EFL graded readers and accompanying back-up materials, but can provide tests for the placement of a student into a reading scheme and for the formal measurement of a student's progress within the scheme.

For future understanding of the rationale behind the test design, it is important to note here that EPER assigns each EFL graded reader to one of eight EPER levels, overriding the publishers' own level classification. The EPER levels are: G, F, E, D, C, B, A and X - G being the easiest and X the most difficult.

The purpose of a test is to decide which EPER level a student should "rightly" be reading at. From the user's point of view, this simply means that each possible raw score on the test should be interpretable in terms of an EPER reading level.

Until now, EPER has used a pair of standardized cloze tests to decide a student's initial reading level and, with subsequent administerings of the tests, to see how far up the EPER ladder of levels a student has moved. These tests have the advantages of being easy to administer and easy to mark, and the scores are immediately convertible to EPER reading levels using the scores conversion tables made available to EPER's customers.

However, there are also major disadvantages to these tests. One point of dissatisfaction is that, since the cloze tests are here being offered as tests of extensive reading competence rather than as tests of general language proficiency (which is the more widely accepted use of cloze tests), they do have extremely little face validity. Many teachers, students and administrators do not see any immediately visible connection between a cloze test score and extensive reading performance. The point about face validity is not whether the test is or is not an accurate measure of what it purports to measure, but whether it is seen to be such by the users. The cloze tests' lack of face validity in this context has proven to be unsettling for some of EPER's clients.

A second argument against the cloze tests is that they can be a very traumatic experience for lower level students. The two tests, Cloze A and Cloze B, are designed to measure all levels of proficiency. Although Cloze A is split into first half (the easier part) and second half (more difficult), Cloze B is given to students in its entirety. This means that the lower level students are faced with a test over half of which they have no hope of being able to complete. The weakest students may have to leave nine tenths of the answer-sheet blank. Unless Cloze A is to be given repeatedly, all the students in a reading programme which uses the tests will face Cloze B at some time in their extensive reading careers. In any case, Cloze A also, despite its division into two parts, will give rise to the same kind of situation, with some students effectively only able to perform on a quarter or a half of the test they have been given. This is not really acceptable for many teachers. It can also affect a test's reliability as a true measure of student performance, since a student who experiences a sense of defeat before even putting pen to answer-sheet will quite possibly perform a lot worse than he otherwise might have done.

These two concerns have provided a large part of the impetus for the development of the new extensive reading test.

2. Structure of the test

As it stands at the moment, the new test consists of two component parts - a reading comprehension paper and a separate discrete-item multi-choice vocabulary test.

2.1 Vocabulary test

Whereas the reading comprehension paper is stratified (and will be discussed later), the vocabulary test is a single unit and the same test is given to all students at every level. The vocabulary test consists of 70 items. Items are graded, so that the easier items are at the beginning, with the test becoming progressively more difficult.

The vocabulary test is subject to some of the same criticisms as the cloze tests. Firstly, with items designed to measure performance from beginner to near native-speaker level, it may be rather off-putting for the weaker students. Secondly, it has less face validity than the reading comprehension component which looks like a reading test, (although it has arguably more face validity than the cloze tests, since at least vocabulary acquisition is fairly easy to associate with extensive reading).

Two further arguments against the vocabulary test are found in the questions of construct and content validity. As Hatch and Farhady put it, the '...problem in construct validity is whether our test items really comprise the construct...' (Hatch and Farhady 1982: 252). In this particular case the problem would be whether a multi-choice vocabulary test can really be a measure of reading competence. Most of us would intuitively agree that an increase in a student's receptive vocabulary will be one of the gains from reading extensively, but most of us would also agree that the reading process must involve far more than receptive vocabulary knowledge. At best then, the vocabulary test can only comprise part of the extensive reading construct.

As for content validity - defined by Hatch and Farhady as '...the extent to which a test measures a representative sample of the subject matter content...' (Hatch and Farhady 1982: 251) - one problem might be simply that the vocabulary test is too short. Given that the test must discriminate eight different levels, and there are 70 items in the test, discrimination between any two adjacent levels will largely depend on a student's response to fewer than ten items. Are nine responses really enough to gauge a student's vocabulary knowledge? Even under the impossible circumstances of each item being perfectly reliable and each student and each marker performing perfectly reliably and there being no overlap between levels, it is not likely that nine vocabulary items could be properly representative of the vocabulary in the readers at any given EPER level.

To increase its content validity the vocabulary test would need to be made much longer - but then it would be even more like Cloze B, in that the longer the test, the greater the number of items which are inaccessible to the lower level students. One answer to this would of course be to split the test into two halves or stratify it even further into anything up to eight levels. The question of construct validity would however remain.

The present recommendation to the test's users in Hong Kong (which is where the test was piloted and which is, at the moment, the only place where it is used) is, where possible, to use scores from the reading comprehension component rather than from the vocabulary test component to determine a student's EPER reading level. It is even possible that the final version of the extensive reading test will be reduced to the reading comprehension component only. Quite apart from the other considerations, the more compact a test, the more administratively practical it is - particularly important where thousands of testees are to be involved. Whether the vocabulary test will eventually be removed, or whether it will remain in some kind of complementary capacity to the reading comprehension component will depend on the results of the follow-up research into the test's use in Hong Kong which is planned for 1993.

2.2 Reading comprehension papers

Whereas the vocabulary test is a single unit intended to accommodate all levels, the reading comprehension paper is in fact eight separate reading comprehension papers - one for each EPER reading level. The only feasible way to have one single reading comprehension paper for all levels would be to have one paper made up of increasingly difficult comprehension passages. Honouring the distinction between "extensive" and "intensive" reading as best we can (extremely difficult when attempting to write a test since testing is by nature an intensive activity, unlike continuous assessment), and having rather longer than usual passages, this could have resulted in a three- to four-hour reading comprehension paper. Again, the weaker students would be demoralised and class time would be wasted as they sat for three hours in front of an impossible task. Conversely, the stronger students would be wasting their time on material far too easy for them, and marking would take longer. Moreover, the students would end up sitting the whole composite paper again and again. In view of all this, it seemed better to have eight separate papers.

The eight papers were then grouped as four pairs of adjacent papers. Each student takes a pair of papers - the combined scores on two papers giving a more reliable result than a score from one paper only. The decision as to which pair of papers a student will take ultimately rests with the teacher and will depend upon the student's current reading level. However, if the student gets above a certain combined score on a pair of reading comprehension papers, then the student should be given the two reading comprehension papers immediately above. Likewise if a student obtains below a certain score then he should be given two easier reading comprehension papers. The middle range of scores from any two paired papers is divided into two bands, each band of scores pertaining to one EPER reading level. Thus a student need not take more than four papers (two pairs), but the vast majority of students will need to take only one pair of papers. This will not only save everyone's time, but will cut down on student frustration arising from being faced with material which is far too easy or far too difficult.

So far as we know, the Hong Kong administration is very pleased with the new more user-friendly reading comprehension test, which also has more face validity than the old cloze tests. It is also my personal belief that teachers in Hong Kong will feel more personally involved in the test than they did with the old cloze tests, since their initial judgement on a student's EPER reading level is what is used to route the student towards the appropriate pair of tests. Teachers are thus asked to make a professional contribution to the testing machinery, something they were not asked to do with the cloze tests where all students automatically took the same test. I believe that this professional involvement will have a favourable effect on teachers' attitudes towards the test. (Whether this is indeed the case will be researched in the 1993 follow-up study.)

3 Test validation

Both the vocabulary test and the reading comprehension papers were piloted in Hong Kong, and the results analyzed at Edinburgh.

3.1 Validation of the vocabulary test

Analysis of the vocabulary test was very straightforward. A Rasch analysis gave each item a difficulty estimate and these difficulty estimates were then used to convert each possible raw score on the vocabulary test to a student ability estimate. A scale of abilities was then devised with eight student ability bands - each band corresponding to one EPER reading level. The top and bottom cut-off points for each band (and hence for each EPER reading level) were decided through post-hoc comparison with known cloze scores (each student in the pilot already having been assigned an EPER reading level on the basis of a cloze score). That is to say that, for example, the typical ability estimates for students already assigned to EPER reading level B through their cloze scores would be used as the student ability estimates for band B on the vocabulary test and students demonstrating those same abilities on the vocabulary test would be assigned to level B. In other words, the vocabulary test was validated and stratified against the cloze tests.

This obviously raises the question of whether such a validation is really tenable, given that a cloze test and a multi-choice vocabulary test are two quite different testing beasts and may be measuring quite different things. The correlation between the cloze scores and the vocabulary test scores was however .8 and it will be interesting to see how well the comparison between the two test-types holds in practice.

3.2 Validation of the reading comprehension papers

Validation of the reading comprehension component was a little less straightforward. Although the eight reading comprehension papers were conceived in such a way that

87

putting them all together would provide a complete test of reading comprehension ability from beginners to very advanced, no student sat the complete test. Thus a correlation between reading comprehension scores and existing cloze scores or the vocabulary test scores would have been nonsensical. (For example a student sitting the two lowest level reading comprehension papers may well have obtained a very high score on these, but - being at a lower level - would have obtained a very low cloze score. A student who obtained a very high score on the cloze, but who sat the highest level reading comprehension papers, might have a lower reading comprehension score than the low level student.)

Eight separate correlations - i.e. between the scores for each separate level of the reading comprehension test and cloze scores - might have made more sense. However there was not a wide enough range of scores at certain levels of the reading comprehension papers for a satisfactory correlation to be made. Nor was there a large enough number of scores at certain levels. Although the total number of students taking part in the pilot was over 200, the number for each level was sometimes less than thirty. (Hatch and Lazaraton (1991: 550) recommend a minimum N of 35.)

The other problem with the series of reading comprehension papers was how to interpret scores across papers. To give an example: if a student obtained a score of 30 on the lowest level pair of papers, then six months later obtained a score of 20 on the immediately higher pair of papers, how would a teacher know how much progress, if any, had been made by that student, given that the papers are pitched at different levels?

The obvious answer is regression, which would compute equivalent scores on different pairs of papers, but, as with correlation, the scores on the reading comprehension papers did not meet the technical requirements for regression. (Notably the numbers at certain levels were again too low, and the ranges of scores at some levels were not wide enough to establish a correlation - a high correlation between two groups of scores being a pre-requirement of regression.)

The procedure in fact followed was to place all eight reading comprehension papers consecutively one after the other to form one long test and to perform a Rasch analysis on all the items at the same time. This was possible because, although no student took more than two adjacent levels of reading comprehension papers, there was linking throughout the complete series of papers. That is to say that, for example, some students took the levels G and F papers and others took the levels F and E papers. The results of the F papers would provide the link between the G and E papers necessary to standardize item difficulty estimates across the three levels on to a common scale. With linking throughout the eight levels, item difficulty estimates would be standardized to a common scale throughout the whole test.

One problem with this, however, is that a reading comprehension test is not an ideal candidate for Rasch analysis, since the Rasch model assumes that all items are discrete, and that the chances of an item being answered correctly or incorrectly are unaffected by answers on preceding items. This is clearly not always the case in a reading comprehension test, although it does depend on the particular items and attempts were made at the item-writing stage to produce independent items. Again, it will be interesting to find out how well the Rasch item difficulty estimates for the reading comprehension papers perform in practice.

The Rasch item difficulty estimates were then used to produce student ability estimates from raw scores on pairs of papers. To give an example: a combined score of 30 on the levels F and G papers would show an ability rating of -0.2, which is the same ability rating as a combined score of 18 on the levels D and E papers.

The next logical step would be a scores conversion chart. However there are many factors affecting the reliability of such ability estimates, and in order to avoid any spurious accuracy the scores were not reported as individual scores and equivalent scores on different pairs of papers, but again a bands system was used. The same procedure was used to decide upper and lower ability estimates for each band as was used for the vocabulary test bands. The ability estimates of students known to be reading at a certain level (allocated to that level because of their cloze scores) were taken as the prescriptive ability estimates for that EPER reading level.

Thus the reading comprehension component, like the vocabulary component, was validated against the cloze tests. The same question arises - how legitimate was this procedure? The follow-up study will hopefully give some indication.

4. Conclusion

One of the main problems in constructing a test for all levels is the avoidance of material which is much too easy or much too difficult for some of the students. The answer to this in our case was a stratified test divided into eight levels. However, this stratification brought with it a number of problems at the validation stage, particularly how to standardize scores at different levels on to a common scale. Several standard statistical procedures (correlation and regression) could not be used because of the low numbers of students who took part in the pilot at each level. Our answer was to treat the eight papers as one continuous test and then perform a Rasch analysis. We then used Rasch item difficulty estimates and student ability estimates to compute equivalent scores across levels.

An extensive follow-up study for the test is planned for 1993. How well this validation procedure works in practice will then be investigated.

References

Hatch E. and H. Farhady. 1982. Research Design and Statistics for Applied Linguistics. Rowley, Massachusetts: Newbury House.

Hatch E. and A. Lazaraton. 1991. The Research Manual: Design and Statistics for Applied Linguistics. Rowley, Massachusetts: Newbury House.

QUESTIONS IN LECTURES: OPPORTUNITIES OR OBSTACLES ?

Tony Lynch (IALS)

Abstract

One of the consequences of the rise in numbers of nonnative students at British universities is an increased risk that lecturers will fail to make themselves adequately understood to heterogeneous audiences. Although listeners may be invited to ask questions, there are linguistic, psychological and sociocultural pressures on nonnative students that can deter them from doing so. This paper discusses the nature of those pressures on would-be questioners and suggests ways in which teaching staff could make the asking and answering of questions less inhibiting. This would bring benefits in terms of the accessibility of lectures to both nonnative and native listeners.

1. Introduction

As the numbers of students undertaking higher education outside their home country increase, the institutions receiving them are having to devise ways of catering for a student population that includes a substantial proportion of nonnative speakers (NNSs). Most of the effort in that direction is focussed on providing language and study skills tuition for incoming students whose linguistic competence is thought to place them at risk of academic failure. Such tuition may be basically preventive or remedial, taking the form of pre-sessional courses preparing students for entry into the institution or of in-session classes run after the students' main course has started.

For many NNS students the principal problem encountered at the start of their academic course is the difficulty of understanding lectures. Comprehension of the local spoken form of the language is of course a common problem for anyone newly arrived in a foreign country, but the comprehension of lectures raises the additional problem that students' ability to understand, process and note down orally presented information in the first few weeks of the academic session can strongly influence their subsequent performance in written assignments and examinations.

It is for this reason that most English for Academic Purposes (EAP) courses provide a substantial amount of practice in the basic skills of listening and notetaking. In designing the lecture comprehension components of pre-session courses, EAP staff may choose among a range of options: to use published texts (e.g. McDonough 1978; Lynch 1983; Mason 1983; James, Jordan, Matthews and O'Brien 1990); to record material from content lecture courses, normally at other departments of the host institution; or to include guest lectures course given by staff members from the students' future department (Lynch 1984). In-session course designers may also adopt a team-teaching model such as that described by Dudley-Evans and Johns (1981), in which the subject

ISSN 0959-2253

lecturer and language tutor collaborate to provide a listening/language component to supplement an existing content lecture course.

A considerable amount of effort and thought has therefore been devoted to ways of helping NNS listeners to tune in to the characteristic patterns of lecture discourse. Much less attention has been paid, at least in Britain, to providing assistance to the lecturers. In the United States, universities' increasing use of NNS graduate researchers as teaching assistants on undergraduate courses has led to demands (from both the NNS assistants and their students) for programmes to improve the assistants' speaking skills, and the scale of what is often referred to as the 'foreign T.A. problem' can be gauged by the growth in the related literature (e.g. Bailey, Pialorsi and Zukowski-Faust 1984; Rounds 1987; Byrd, Constantinides and Pennington 1989; Pica, Barnes and Finger 1990).

However, there seems as yet to be no published work on the possible implications of the growing numbers of international students in lecture audiences for the way native speaker (NS) teaching staff package and deliver the content of their lectures or, more specifically, for (re-)training programmes to encourage adaptation to a changing student population. In this paper I discuss what is known about one specific area of lecture discourse - questions from the audience - and how that might provide a starting point for (re)training programmes for lecturers to teach multinational classes. The discussion will draw on two main sources: research into NS/NNS interaction and the lecturing methodology literature.

The reason for concentrating on the issue of questions is simply that the answering (and asking) of questions in lectures is difficult enough, even when both speaker and questioner are operating in their own language (Gibbs, Habeshaw and Habeshaw 1987). The additional problems that can arise when the would-be questioner is a second language user make communication even more complex. Both parties may be reluctant to exploit the potential benefits of audience questions. For the lecturer, a point raised by a student may take the discourse into a side-track (or even lead to a complete derailment). For the students, there are other problems. Apart from the burden of public performance involved in asking a question in front of a large audience, those who do venture a question run the risk of being considered (in the British student culture) 'stupid, attention seekers or creeps' (Gibbs et al. 1987: 155). These difficulties are of course compounded when the questioner is a NNS student by their greater unfamiliarity with the language, the academic culture, or both (Ballard 1984; Dunkel and Davy 1989). We will come back to this question of sociocultural adjustment shortly.

2. Native/nonnative interaction research

Studies of the characteristics of communication between native speakers (NSs) and nonnative speakers (NNSs) have established the importance of the role played by questions in the negotiation of meaning (e.g. Hatch 1978; Long 1981, 1983) and have resulted in the development of a taxonomy of 'listener queries' (Rost 1990) geared to the resolution of ambiguity: the 'clarification request', the 'confirmation check', the 'comprehension check', and so on.

Research into the particular case of NS/NNS interaction in second language classrooms (extensively reviewed by Chaudron 1988) has highlighted two potential benefits to be gained by NNS learners' deployment of such questions. First, these modifications of interaction have been shown to be more frequent and more consistent than adjustments of input, or language form, (Long 1981) and also to be more likely to enhance the comprehension of NNS learners (Pica, Young and Doughty 1987). Second, a number of authors have argued that a realignment of discourse roles is necessary for the development of a fuller second language competence than can be achieved if NNS learners are restricted to a passive/responsive classroom role (Pica 1987, van Lier 1988, Lynch 1991).

However, one of the complicating factors in any attempt to encourage learner-to-teacher questions is the expectations that the participants bring to the classroom. Many NNS learners will expect the teacher to fulfil the roles of possessor of knowledge and of authority figure that they are familiar with from home, rather than those of informant and facilitator, which may be assumed in the teacher's own approach: 'given the unequal relationships of teacher and student established by the design and organisation of classroom activities, students may begin to feel that their clarification requests and confirmation checks will be perceived as challenges to the knowledge and professional experience of the teacher' (Pica 1987: 12). When the focus shifts to the lecture theatre, as opposed to the L2 classroom, where the lecturer carries the additional authority of content specialist, one might reasonably assume that such NNS listeners will be even more reluctant to intervene and ask questions. Conversely, NNS students coming from an educational culture in which students can and do interrupt lecturers at any point in a lecture may do so more than is expected in the British context, even if they attempt to restrict their interruptions, having recognised that the cultural norms are different.

Turning now to research into NS/NNS lecture discourse, we find that a number of studies have established ways in which lecturers can help NNS members of their audience by modifying their spoken discourse. Linguistically, this includes speaking at a slower pace with clearer articulation and with a greater degree of verbal and visual redundancy (Chaudron 1983; Wesche and Ready 1985; Olsen and Huckin 1990). Rhetorically, more overt signalling of discourse structure and development and of key points appears also to enhance NNS comprehension (Chaudron and Richards 1986). But, as Wesche and Ready have noted, crucial to any discussion of what may help NNS listeners to understand lectures is the extent to which the lecturer is willing to help: native speakers will also vary in their underlying sensitivity to - and even interest in - the comprehensibility of their input to nonnatives (1985: 108).

Olsen and Huckin (1990) argue that what is required for adequate lecture theatre competence is the ability to achieve 'point-driven', rather than 'information-driven', understanding, i.e. that a NNS listener needs to be able to follow the overall development as well as recognise the detail. This conclusion was reached after their discovery that some of the NNS listeners in their study failed despite adequate English, which reinforces the point that competence and ease of lecture comprehension and notetaking is not simply a question of language ability (cf. Dunkel and Davy 1989).

3. Lecturing methodology

It is revealing that in much of the British literature on lecture methodology (e.g. Costin 1962; Bligh, Ebrahim, Jaques and Piper 1975; University Teaching Methods Unit 1976; Curzon 1980), the word 'question' is used exclusively to refer to questions asked of the audience by the lecturer, rather than vice versa, with all that implies about the relative statuses of asker and asked. Expressed in the terms used in NS/NNS research, 'question' in this field means *comprehension check* rather than *clarification request*. However, one exception to this general trend is the work of George Brown (Brown 1978; Brown and Bakhtar 1983; Brown and Atkins 1988), who recognises that question-asking in lectures is a communicative activity in which there can be an advantage in the listener, as well as the speaker, taking the initiative. Brown and Bakhtar (1983) include the following points in their widely cited set of recommendations to new lecturers:

(1) Speak loudly and clearly ... don't go too fast.

(2) Plan, prepare, structure every lecture.

(3) Make it understandable - explain, emphasise, recap, repeat and summarise main points and relate to current examples and applications.

(4) Watch out for reaction and feedback, invite questions and ask questions, encourage participation, involve your audience.

Item 4 in what may appear to be an unexceptional list in fact contains the potential for revolutionary change. Consider what might happen if lecturers did invite questions from the audience. For many lecturers, it would at the very least create 'tension between the teacher's authority (expressed in his control over content) and his aim of making himself receptive to feedback' (Startup 1979: 29). On the similar issue of allowing questions in business presentations, Jay has written that 'The power of questions to help a presentation is less than their power to damage it' (Jay 1971: 67).

However, Brown's call for lecturers to encourage audience participation through questions has been echoed by other writers, who provide practical recommendations as to how this might work: Cannon (1988) suggests avoiding the stress of public questioning by asking the students to make a note of any questions on slips of paper, for the lecturer to collect in and choose from when deciding which points to respond to. Gibbs et al. (1987) propose group-based discussion of points that students want clarified; this would allow them also to decide on a suitable wording for the question, again relieving any one student of the burden of individual performance.

4. Sociopragmatics of questions

An essential preliminary in training lecturers in techniques of dealing with mixed audiences is that they should be made aware of the possible sociocultural problems faced by NNS students entering university. It should be stressed that these are not restricted to second language speakers; however, the degree of unfamiliarity and alienation is likely to be more severe for NNS students. Ballard (1984), investigating the adaptation problems of NNS students entering Australian university, coined the phrase 'double cultural shift' to describe the situation of the second language/culture learner moving both from secondary school to university (or from undergraduate to postgraduate course), and also from home to alien culture, with different norms of authority, personal responsibility and so on. Texts dealing with sociocultural aspects of study abroad, such as the collection edited by Adams, Heaton and Howarth (1991), would provide an appropriate perspective on some of the major issues facing NNS university entrants.

Similarly, Olsen and Huck[illegible] (1990), Shaw and Bailey (1990) and Strodt-Lopez (1991) have stressed the need to 'initiate' NNS listeners into the local expectations of lecture theatre behaviour (by lecturer and by students). This is something that has also been recommended in the general methodology literature (e.g. Gibbs et al. 1987; Ellington n.d.) but would be of additional value in the case of lecture audiences with NNS members.

Earlier I referred to Pica's (1987) argument that second language learners may be unwilling to ask the language teacher to repeat or clarify, for fear that such queries may be taken as a slight on the competence or authority of the teacher. The extent to which NNSs' perceptions of the pragmatics of questioning can vary is illustrated by two classroom incidents from EAP courses at IALS. In one case I was working with a group of Indonesian tax officials and had dealt rather unsuccessfully with a request for explanation of a grammar point. I thought I should check that the learners had understood my explanation; the following exchange then took place between the senior student, who usually acted as spokesman for the group, and myself:

T: Would you like to ask any questions about that?
S: *(immediately)* No questions.
T: What about the others?
S: They have no questions.
T: But how do you know the others don't have any questions?
S: Because you are a good teacher.

At the other end of the spectrum was the reaction of a group of Swedish lecturers in science and technology who attended a short course at IALS prior to a period of attachment in various departments at Edinburgh, as part of a scheme to prepare them for teaching international groups of students through English in Sweden. While we were discussing the issue of handling questions in lectures, I asked when they preferred students to ask questions. They seemed perplexed and asked what I meant. When I repeated my question, one said, 'Well, you answer a question when it's asked, don't you?' and the others nodded. Clearly, for this Swedish group, a lecture seemed to be more informal and more conversation-like (at least, with more turn-taking) than would be the norm in Britain. Confirmation of this came when we met after they had spent a week in University of Edinburgh departments and they talked of how surprised they had been by the total absence of questions from students. One had even asked a British student whether he had understood everything in the lecture and was told that he had not; on then asking why the student had not asked for clarification, he was told, 'I go and look it up in the library'.

5. Implications for lecturer training

One important element in any training programme would be to advise lecturers to take time at the start of the lecture course to make clear their personal preferences for the form and timing of audience questions: whether they can be asked during the lecture or afterwards; whether queries will be discussed in plenary, or whether it is up to the individual student to ask the lecturer at the end of the lecture, or by making an individual appointment. Seen in black and white, (as here), such advice may appear trivial, but the evidence is that scene-setting at the beginning of academic courses is rare; as Shaw and Bailey (1990) have shown, NS students are left to work out each of their lecturers' individual preferences about matters such as question-handling on the basis of hints during the first few sessions of a lecture course. All the more reason, then, for setting out the ground rules explicitly for an international class.

A second element would be the suggestion that lecturers should schedule in two or three 'question pauses' - short breaks in their presentation during which students would be free to raise queries about what has been said up to that point. The advantage of clearly signalling 'time for questions' would be firstly to allow listeners time to review what they have just heard and to formulate questions, and secondly to remove the necessity of bidding for a turn while the lecturer is speaking. Such question pauses, providing an overtly marked space for clarification requests, could do a great deal to assist NNS students to take the initiative in raising points they need to have explained.

Thirdly, lecturers could be given practice in negotiating meaning with NNS questioners. It can be difficult to understand audience queries - whether at the level of intelligibility, comprehensibility or interpretability (Smith and Nelson 1985) - and that problem can become more acute if the questioner comes from a society where it is customary to make the act of questioning more acceptable by expressing the question obliquely. In particular, practice in repeating or rephrasing audience questions - cf. the *confirmation check* of NS/NNS interaction research - should also feature in a lecturer training programme. Seminar skills materials designed for NNS students (e.g. Lynch and Anderson 1992) are one potential source of exercises in appropriate negotiation practice.

6. Conclusion

Much of the work done on pre-sessional courses for incoming students is based on the assumption that a well-planned and well-executed preparatory course can prevent problems arising - in the specific case of lecture comprehension, by improving learners' listening skills to the point where they will understand adequately. However, we cannot guarantee that they will encounter no problems; indeed, since we know that native listeners also experience difficulties (even if less frequent and less marked), we should expect problems to arise. Two practical training approaches would help to reduce the problems: the first would be to provide NNS learners with practice in identifying uncertainties and formulating concise and transparent questions; the second, discussed in this paper, would be to help lecturers unfamiliar with the needs of an international audience to find ways of dealing with comprehension problems when they arise.

The fact that many studies of L2 lecture comprehension characterise the spoken information as input highlights a general imbalance in the way the lecture has been represented as a communicative event, with the emphasis on the transmission of information to an audience. Although this close analysis on what lecturers say and do has resulted in an increased awareness of the benefits for comprehensibility of a clearly signalled discourse structure, there is surely also a case for enhancing lecturers' appreciation of the benefits of making lectures more interactive by encouraging clarifying questions.

Higher education institutions will continue to run study skills courses that develop NNS students' listening comprehension and notetaking skills, but we need also to assist lecturers to cope better with the demands of teaching international classes. Training which emphasises some of the potentially helpful strategies in NS/NNS communication, such as the questionin discussed in this paper, should make lectures more successful communicative events for native and nonnative listeners alike.

References

Adams P, B. Heaton and P. Howarth. 1991. Socio-Cultural Issues in English for Academic Purposes. Review of English Language Teaching 1/2. Modern English Publications / British Council.

Bailey K., J. Pialorsi and J. Zukowski-Faust (eds). 1984. Foreign Teaching Assistants in U.S. Universities. Washington, DC: National Association for Student Affairs.

Ballard B. 1984. 'Improving student writing: an integrated approach to cultural adjustment' in R. Williams. J. Swales and J. Kirkman (eds). 1984. Common Ground: Shared Interests in ESP and Communication Studies. ELT Documents Special. Oxford: Pergamon. 43-53.

Bligh D., G. Ebrahim, D. Jaques and D. Piper. 1975. Teaching Students. Exeter: Exeter University Teaching Services.

Brown G.A. 1978. Lecturing and Explaining. London: Methuen.

Brown G.A. and M. Atkins. 1988. Effective Teaching in Higher Education. London: Methuen.

Brown G.A. and M. Bakhtar (eds). 1983. Styles of Lecturing. University of Loughborough: ASTD Publications.

Byrd P., J. Constantinides and M. Pennington. 1989. The Foreign Teaching Assistant's Manual. New York: Collier Macmillan.

Cannon R. 1988. Lecturing. Kensington, NSW: Higher Education Research and Development Society of Australia.

Chaudron C. 1983. 'Simplification of input: topic reinstatements and their effects on L2 learners' recognition and recall'. TESOL Quarterly 17/3: 437-58.

Chaudron C. 1988. Second Language Classrooms. Cambridge: Cambridge University Press.

Chaudron C. and J. Richards. 1986. 'The effect of discourse markers on the comprehension of lectures'. Applied Linguistics 7/2: 113-27.

Costin F. 1972. 'Lecturing versus other methods of teaching: a review of research'. British Journal of Educational Technology 3: 1-31.

Curzon L. 1980. Teaching in Further Education. London: Cassell.

Dudley-Evans T. and T. Johns. 1981. 'A team teaching approach to lecture comprehension for overseas students' in The Teaching of Listening Comprehension. London: British Council. 30-46.

Dunkel P. and S. Davy. 1989. 'The heuristic of lecture notetaking: perceptions of American and international students regarding the value and practice of notetaking'. English for Specific Purposes 8/1: 33-50.

Ellington H. n.d.. Some Hints on How to Be an Effective Lecturer. Teaching and Learning in Higher Education 6. Scottish Central Institutions Committee for Education Development.

Gibbs G., S. Habeshaw and T. Habeshaw. 1987. 53 Interesting Things to Do in Your Lectures. Bristol, England: Technical and Educational Services.

Hatch E. 1978. 'Discourse analysis in second language acquisition' in E. Hatch (ed.). 1978. Second Language Acquisition. Rowley, MA: Newbury House.

James K., R. Jordan, A. Matthews and J. O'Brien. 1990. Listening Comprehension and Notetaking Course. (2nd edition). London: Collins.

Jay A. 1971. Effective Presentation. London: British Institute of Management Foundation.

Long M. 1981. 'Input, interaction and second language acquisition' in H. Winitz (ed.). 1981. The Comprehension Approach to Foreign Language Instruction. Rowley, MA: Newbury House.

Long M. 1983. 'Native/nonnative conversation and the negotiation of comprehensible input'. Applied Linguistics 4/2: 126-41.

Lynch T. 1983. Study Listening. Cambridge: Cambridge University Press.

Lynch T. 1984. 'Listening comprehension in English for specific purposes: Integrating specialist lectures into pre-sessional courses'. English Language Research Journal 5: 10-27.

Lynch T. 1991. 'Questioning roles in the classroom'. English Language Teaching Journal 45/3: 201-10.

Lynch T. and K. Anderson. 1992. Study Speaking. Cambridge: Cambridge University Press.

Mason A. 1983. Understanding Academic Lectures. Englewood Cliffs, NJ: Prentice Hall.

Olsen L. and T. Huckin. 1990. 'Point-driven understanding in engineering lecture comprehension'. English for Specific Purposes 9: 33-47.

Pica T. 1987. 'Second-language acquisition, social interaction, and the classroom'. Applied Linguistics 8/1: 3-21.

Pica T., G. Barnes and A. Finger. 1990. Teaching Matters: Skills and Strategies for International Teaching Assistants. Rowley, MA: Newbury House.

Pica T., R. Young and C. Doughty. 1987. 'The impact of interaction on comprehension'. TESOL Quarterly 21/4: 737-58.

Rost M. 1990. Listening in Language Learning. London: Longman.

Rounds P. 1987. 'Characterizing successful classroom discourse for NNS teaching assistant training'. TESOL Quarterly 21/4: 643-71.

Shaw P. and K. Bailey. 1990. 'Cultural differences in academic settings' in R. Scarcella, E. Andersen and S. Krashen (eds). 1990. Developing Communicative Competence in a Second Language. Rowley, MA: Newbury House. 317-28.

Smith L. and C. Nelson. 1985. 'International intelligibility of English: directions and resources'. World Englishes 4/3: 333-42.

Startup R. 1979. The University Teacher and his World: a Sociological and Educational Study. Aldershot: Gower.

Strodt-Lopez B. 1991. 'Tying it all in': asides in university lectures'. Applied Linguistics 12/2: 117-40.

University Teaching Methods Unit. 1976. Improving Teaching in Higher Education. London: University of London.

van Lier L. 1988. The Classroom and the Language Learner. London: Longman.

Wesche M. and D. Ready. 1985. 'Foreigner talk in the university classroom' in S. Gass and C. Madden (eds). 1985. Input in Second Language Acquisition. Rowley, MA: Newbury House. 89-114.

98

CAN APPLIED LINGUISTS DO ETHNOGRAPHIC INTERVIEWS?

Brian Parkinson (IALS)

Abstract

Eight subjects who had used 'study packs' in their learning of French and Italian were interviewed by colleagues of the teachers who wrote them. This article presents, not the findings of the interviews, but an analysis of attempts at an 'ethnographic' interviewing strategy, entailing inter alia an open-ended approach and adoption of an 'outsider' role. A coding system designed to measure 'ethnographicity', with sample codings and descriptive statistics, is presented, together with subjective analyses of sample interviews. The surprising and highly provisional conclusion is that 'insider' interviewers can sometimes achieve similar results to ethnographers, but by rather different means.

1. Introduction

This article is one product of a research and development project at IALS concerned with distance learning study packs for intermediate students of French and Italian. The other products are the study packs themselves, in French (Mulphin 1991) and Italian (Dawson and Peyronel 1991), and a summary - not intended for publication - of learners' attitudes to these materials and to more general issues of distance learning and self-study (Howard 1992). This article is based on the same interviews as Howard's summary, but it deals, not with primary research findings on learner attitude etc., but with secondary or 'meta-research' issues, explained further below, concerning interviewer behaviour and the subjects' perception of the interviewers.

The study packs consisted of an audio-tape with several foreign-language interviews and a written booklet with a variety of exercises based on this material plus some reading-based exercises. They were supported by a marking service: students could send in and receive correction and feedback on written exercises, and also (though this was rarely done in practice) a taped oral composition. They also filled in a 'diary page' describing when, where and how they used the materials.

Materials and feedback were free to the students, as this was a pilot version to be revised in the light of their comments and performance. The materials were written from January to June 1991 and piloted from April to August 1991. It was not pure 'distance learning' as many students continued to come to classes, but their distance work was not normally discussed in class, only written feedback being given in envelopes. Other students were not attending class during the pilot period. A telephone tutoring service was offered to supplement written feedback but was scarcely used.

EDINBURGH WORKING PAPERS IN APPLIED LINGUISTICS, number 4 (1993) ISSN 0959-2253

Evaluation of the pilot was by two kinds of interview, 'short interviews' and 'long interviews'; assignment of students to each type was random as far as possible but constrained by student availability for long interviews. All interviews were in English. Intended length of the short interviews was about 10 to 15 minutes, long interviews 45 to 60 minutes, but this varied in practice and some 'short' interviews were longer than some 'long' ones. The short interviews were conducted by the French/Italian teachers who wrote the material and were intended to provide formative information for any necessary revision; the long interviews were conducted by Ron Howard in the role of 'distance learning coordinator' and Brian Parkinson in the role of 'project research adviser', and were intended to be broadly 'ethnographic', to illuminate more general issues of how distance learning materials are perceived and used by students. It was hoped that by presenting ourselves (RH/BP) as 'outsiders' - we were not the writers, though we had advised them - we would get an, in some sense, truer picture of such matters.

2. The ethnographic interview

The ethnographic interview (see e.g. Spradley 1979) is intended as a solution to a well-known methodological problem of interviewing: that the interviewers set the agenda, the interviewees tell them what they think they want to hear, and so there is no real insight into the life-world of the interviewees. Ethnographic interviewers try to treat their subjects as teachers and themselves as learners, thus gaining insight into the subjects' life-world.

We are not professional ethnographers, our previous experience of ethnography was very limited (BP) or none (RH), and full-scale ethnography is not possible in a single interview, so the approach described below is very much a 'diluted' version of that advocated by Spradley. (It was diluted even further for RH as he had a specific conscious agenda in his role as distance-learning coordinator. I had no conscious agenda, but quite possibly - like any interviewer? - an unconscious one.) Nevertheless, we both tried to direct the interview along generally ethnographic lines, particularly in its early stages. The extent of our success is the main topic of this article.

The ethnographic approach, as we interpreted it and planned to implement it, was as follows:

(i) We would stress that we were not members of the course-writing team, and had only a limited knowledge of the course materials. The interviewees were to treat us as completely ignorant, and teach us, as they would a complete outsider (e.g. a newspaper reporter), about the materials and how they had used them.

(ii) We did not, except incidentally, want specific information on how to revise these study packs. Rather, we were interested in the general idea of distance learning of languages and 'what it means to you'.

(iii) We did not have a pre-set schedule of questions, which we had to get through. Instead, we would encourage them to 'set your own agenda' and say whatever they considered worth saying about the materials.

(iv) We did, however, have certain guidelines. We would be interested to hear about:

(a) their language learning in general, including objectives

(b) the place of distance learning, and of the study packs, within that learning

(c) a description of the materials as well as a reaction to them

(d) perception of the intended purpose, and success or otherwise, of each general kind of element or activity in the materials - they were to say what the different kinds were.

(v) Where we - especially RH - had a specific 'agenda' of questions we would try to keep these to the end of the interview, and to keep the first and longer part maximally 'ethnographic'.

3. The methodological problem

In the examples described by Spradley (and others, going back to Malinowski 1922) ethnography is possible because the interviewers have a lot to learn - they genuinely do not know about the terminology and customs of the community studied.

We too have a lot to learn - we genuinely do not know how ordinary people learn languages - but our task is more difficult in a way because we are in danger of being perceived, and even of perceiving ourselves, as 'experts'. Even if not intimately acquainted with the materials being discussed, we are professionals in the field: the interviewees are likely to know this, and if we try to disclaim or minimise our experience and expertise we are likely to be perceived as, and to feel, dishonest, thus distorting the interview. Does this mean that one cannot do ethnographic interviews in an area close to one's own specialisation? Or is there in practice no problem, with interviewees able to talk exactly as to someone outside the field, and interviewers able to adopt a 'naïve' perspective?

I did not expect to answer such a big question from such a slender empirical base: I hoped merely to offer some illuminative examples, simple statistics and insights which might help future workers in this area.

4. Data and data analysis

The data consists of the transcripts of eight 'long interviews' (see above), four conducted by RH and four by BP.

I decided to analyse these in the following way:

(i) To generate a list of numbered categories of utterance which I considered relevant to the question of whether the interviewer was adopting an ethnographic perspective, a language expert perspective, or some other perspective, and whether the interviewee was perceiving him as expert, outsider or in some other way. The system would not categorise all utterances but only those, probably a minority, relevant to this issue.

(ii) To test the reliability of these categories by asking colleagues from outside the project to assign a random sample of utterances to one (or none) of them.

(iii) To generate a profile of each of the eight interviews by coding the appearance of each of the numbered categories in each turn of the interview.

(iv) To record my own subjective impressions, and if possible also colleagues' impressions, of the degree of 'ethnographicity' of each interview, and to compare these with the picture given by the numbered profile.

(v) To attempt generalisations on the factors which appear to influence the ethnographicity of an interview, and on how far such information is recoverable from interview data.

This programme is far from complete and the results below are partial and provisional. In particular, no formal reliability trials have yet occurred.

5. Coding system and examples

(Numbers in examples refer to interviews, pages, turns: thus 1, 7, 5 = interview 1, page 7, turn 5. The original categories have been renumbered, and in some cases collapsed, for this presentation.)

5.1 'A' codings

These are categories of interviewer utterance or part-utterance which seem conducive, or intended as conducive, to the interview proceeding along ethnographic lines.

A1 Disclaiming or minimising personal experience of language learning

> *"... what I'm particularly interested in is to try and find out a little bit about what it's like to learn a language by itself. This is actually something I've never attempted to do so I'm completely ignorant ..." (1, 1, 1)*

A2 Disclaiming or minimising knowledge of project

> *"Now it's very important to treat me as completely ignorant, to assume that I know nothing at all about these materials. In fact I don't know a lot." (4, 1, 1)*

A3 Asking for information about materials

> *"Is there a key for this bit?" (5, 6, 6)*

A4 Referring to and inviting expansion of respondent's written comments

> *"And you mostly seem to have studied in periods of between half an hour and one and a half hours ... Do you find this an ideal time? (4, 8, 4 and 4, 8, 6).*

A5 Asking for information on foreign language (vocabulary) in materials

> *"What's 'elenco', 'elenco'?"* (7, 12, 8)

A6 Giving respondent control over course of interview

> *"There's no fixed list of questions. It's really up to you to say whatever you think important."* (3, 1, 1)

A7 Sympathetic echoing

> *"Mmhm. Right. OK. Em ... you said you found it difficult to organise time to do homework for the normal ... normal classes?"* (2, 6, 4)

(5 other, infrequent 'A' categories (A8 to A12) are not separately defined in this report, and are conflated in the table (section 7) as 'A other'.)

5.2 'B' codings

These are categories of interviewer utterance or part-utterance which seem to indicate an (intentional or unintentional) deviation from the ethnographic pattern.

B1 Identifying self with materials writers

> *"Well, the aim is to have six units altogether, and if you can appreciate that one took quite a lot of [time?]"* (1, 10, 16)

B2 Revealing knowledge of or views on content and purpose of materials

> *"It wasn't intended to be a test at all. No. It was to help you."* (1, 4, 16)

B3 Offering to interpret materials or give answers to questions in materials

> *"On the other hand maybe all they wanted you to decide was whether he was a professional man or [...] a labourer."* (5, 10, 14)

B4 Revealing or emphasising own knowledge/experience as language teacher

> *"Now I teach English you see and I do something similar"* (7, 13, 2)

B5 Encouraging the interlocutor, more as teacher-to-learner than as interviewer-to-interviewee

> *"As you say with unit 2, 3 and 4 of course you'll know next time."* (1, 4, 6)

B6 (Giving impression of) going through a checklist of pre-set questions, including follow-ups

> *"And what did you think of the format? The different colours ... ?" (3, 7, 15)*

B7 Leading questions

> *"So sometimes your predictions were right and sometimes they were wrong but even when they were wrong they didn't stop you from learning? (3, 3, 2)*

B8 Completing respondent's unfinished utterances

R *"No, no, no, I, I, I ... was listening ..."*

I *"You were listening specifically to those pros and cons." (1, 8, 1 and 1, 8, 2)*

B9 Imposing own structure on interview

> *"We'll come back to reading later on, but in the listening [...] the first thing you do then is these predictions?" (3, 4, 10)*

(4 other, infrequent 'B' categories (B10 to B13) are not separately defined in this report, and are conflated in the table as 'B other'.)

5.3 'C' codings

These are categories of respondent utterance or part-utterance which seem to indicate acceptance of (or accidental compliance with) an ethnographic pattern for the interview, or at least some aspect of this.

C1 Teaching the interviewer about the materials

> *"Andrea uses different impersonal forms and [you] listen to the tape and find similar ones." (7, 14, 1)*

C2 Informing the interviewer about own performance on the materials (without persistent self-deprecation - cf. D3)

> *"No, I understood the statement. I had no problem with that one at all." (7, 12, 3)*

C3 Criticism of materials

> *"This particular one had a fault in it. It had a bad echo." (4, 1, 2)*

C4 Suggestions for improving materials or methods

> *"The tendency was to look for one of these answers [...] whereas it was a paraphrase of the answer [...] Perhaps ... the question should say 'It may not be the exact answer'." (3, 11, 8 and 3, 11, 10)*

C5 Teaching the interviewer a point of language (vocabulary)

> *"To box. 'Attenuare' means to extenuate basically ..."* (7, 12, 11)

C6 Introducing a new topic (usually some aspect of own learning habits), unprompted or in response to open question

> *"Interestingly enough, I got a [commercial tape course] two months [ago] [...] you're none the wiser, after three months."* (1, 13, 13)

C7 Contradicting interviewer assumption

I *"Half listening and half reading and writing was it?"*

R *"I think it was more listening"* (4, 3, 8 and 4, 4, 1)

(3 other, infrequent 'C' categories (C8 to C10) are not separately defined in this report, and are conflated in the table as 'C other.)

5.4 'D' codings

These are categories of respondent utterance or part-utterance which seem to indicate failure to perceive or non-acceptance of an ethnographic pattern for the interview, or at least some aspect of this.

D1 Use of 'you' or 'your' to refer to the materials or similar

> *"I think it was your ... eh ... introduction was very good."* (1, 3, 9)

D2 Ignoring interviewer's professions of ignorance and statements of interview purpose by answering solely in terms of 'what I liked/disliked about materials' when asked to describe them

I *"Could you tell me then first of all roughly what the study packages are? I mean what kinds of things you find in them, what you do with them?"*

R *"Em ... I found it excellent ..."* (4, 1, 1 and 4, 1, 2)

D3 Extreme self-deprecation in describing performance, i.e. assertions to the effect that 'materials are wonderful, I am stupid'

> *"It's a very good introduction [...] I couldn't have risked my thoughts [...] I found it difficult to talk [...] I was not good at writing down correctly [...] I really felt the study package good [...] Your introduction was very good [...] I got a bit panicky ... I don't think I understood the question [...] I sort of got a bit panicky [...] I was panicking [...] But it was good ... it was good practice."* (Spread over several turns, interview 1, pages 2-4)

(3 other, infrequent 'D' categories (D4 to D6) are not separately defined in this report, and are conflated in the table as 'D other'.)

6. Sample codings

To give a flavour of the analysis, I now give complete sequential codings on two interviews, those discussed in section 8.

Complete turns without any A/B/C/D categories are coded 'I' for interviewer turns, 'R' for respondent turns: this means that nothing obviously 'ethnographic' or 'unethnographic' occurred. Commas separate turns, dashes separate multiple codings within one turn. Page numbers, originally included to assist checking, are retained to give some idea of equal intervals, as turn length varies greatly.

Interview 1

A1, C7, A7, R, A7, C6-D1-C6, A4, C8, B13, C6, A7, C6-C6-C6, (page 2) C4, A7, C4-C6, A1-A3, C1, A8, D3, A2, R, I, R, I, R, I, C2, A7, (page 3) R, B10, R, I, R, A4, R, B11, D1-D3, B3, D1-D3-D3, I, D3, B3, C2, A7, D3, A7, (page 4) C3, B8, C3, A9, D3, B5, D3, A7, R, B12, D4, I, D3, A7, R, B2, (page 5) C4, A7, C4, A7, R, I, R, I, D1, C6, I, R, I, R, I, R, I, R, A6, R, (page 6) I, C6, B8, R, A6, D5, A6, C1, A3, D3, I, R, A6, C2, I, (page 7) C1, A10, C2, A10, C2, I, C2-C2-D3, C2, A10, R, A10, (page 8) C7, B8, R, I, D3, C2, I, R, B3, C2, A7, C2, I, C2, A7, (page 9) C2, A7, C2, A7, C1, A7, C2, A7, C2, A7, C2, I, R, I, R, I, R, A6, (page 10) C1-C2, B5, C2, I, C6, R, I, R, I, B2, D3, I, D1, B1, D1-D3, (page 11) B1, R, I, C2-D6, I, C9, I, C9, I, R, B6, C6, (page 12) B6, R, I, R, B12, C3, B1, R, I, R, I, R, B6, C6, I, C6, B4, R, I, R, A7, (page 13) R, B6, C6-C6, A7, R, B6, R, I, C6, B6, R, B6, R, B6, C6, A7, (page 14) C6-C6-C6, B1, R, B6, R, I, R, B6, R, B6, R, B1, (page 15) R, B6, C6, I, R, B6, R, I, R, B5, R, B1, D1, I, R, I, R, (page 16) B6, R, I, R, I, R, B7, R, B7, D1.

Interview 7

I, R, I, R, A2, (page 2) C1-C2-C8-C2, I, C6, (page 3) A3, C2, I, R, I, C3-D1, A9-B12, R, I, C3, A6, (page 4) R, I, R, A4, R, A4, R, I, R, (page 5) I, C2, I, (inaudible section), (page 6) C2, A7, C2, I, R, I, R, A4, C3, I, C3, I, C3, I, (page 7) R, I, R, A4, R, A4, C2, I, R, (page 8) B7, C2, I, R, I, C2, A6, R, I, C2, I, C2, I, (page 9) R, A3, C1, I, C2, I, C2, B7, C7, I, R, A3, C1, I, R, B7, R, A3, (page10) C1, A3, C1, I, C1, I, C1-C2, I, R, I, (page 11) C1-C2, I, C2, B2, C2, I, R, I, R, I, R, I, C1, I, C2, I, (page 12) C2-C1, B7, C7-C2, I, C2, I, C2, A5, C5, A5, C5, A5, C5, I, R, B3, (page 13) R, B3-B4, R, B3, R, B3, R, B3, R, A7, R, B3, C6, I, C6, I, R, A3, (page 14) C1-C2, I, R, B7, R, I, C2, A3, C1, A3, C1, I, C1, I, R, I, R, I, C1-C2, I, C2, (page 15) I, C6-C4, A3, C6, A3, C6, I, C1-C2, A5, (page 16) C5, A5, C5-C1-C2, A3, C1, A3, C1, I, C2, I, R, I, R, I, (page 17) C1, I, R, I, C1-C2, I, C4, A7, C4, A7, C4, I, (page 18) R, I, R, I, C2, I, C1, A6, R, B6, R, B6, R, I, R, I, R, B6, (page 19) R, B6, R, I, C4, B6, R, I, R.

7. Overall coding statistics

The table below summarises the frequency of the main coding categories across the interviews so far coded. Brief labels are given, but see section 5 for full descriptions and examples.

Interview	**RH Interviews**					**BP Interviews**					**Grand**
Category	**1**	**5**	**6**	**7**	**Total**	**2**	**3**	**4**	**8**	**Total**	**Total**
A1 Disclaim experience	2	2		0	4	0	0	0		0	4
A2 Disclaim knowledge	1	1		1	3	0	3	3		6	9
A3 Ask re. materials	2	12		12	26	0	9	9		18	44
A4 Refer to components	2	3		5	10	0	6	4		10	20
A5 Ask vocabulary	0	0		5	5	0	0	0		0	5
A6 Give control	5	2		3	10	2	4	2		8	18
A7 Echoing	21	4		4	29	5	19	4		28	57
A - Other	6	1		1	8	3	3	2		8	16
A - Total	39	25		31	95	10	44	24		78	173
B1 Identify	6	0		0	6	0	0	1		1	7
B2 Reveal knowledge	2	12		1	15	1	0	0		1	16
B3 Interpret	3	3		6	12	0	0	0		0	12
B4 Reveal experience	1	3		1	5	0	0	0		0	5
B5 Teacher role	3	2		0	5	0	0	0		0	5
B6 Checklist	13	7		5	25	2	7	1		10	35
B7 Leading questions	2	2		5	9	1	7	0		8	17
B8 Completing	3	0		0	3	0	0	0		0	3
B9 Own structure	0	3		0	3	1	5	0		6	9
B - Other	5	0		1	6	0	0	0		0	6
B - Total	38	32		19	89	5	19	2		26	115

C1 Teaching (materials)	5	12		22	39	0	17	10		27	66
C2 Own performance	20	39		32	91	5	9	0		14	105
C3 Criticism	3	3		5	11	3	1	5		9	20
C4 Suggestions	4	1		5	10	2	4	0		6	16
C5 Teaching (vocabulary)	0	0		5	5	0	0	0		0	5
C6 New topic	21	2		6	29	0	9	3		12	41
C7 Contradicting	2	2		2	6	3	0	4		7	13
C - Other	3	0		1	4	0	0	1		1	5
C - Total	58	59		78	195	13	40	23		76	271
D1 Use of 'you'	8	0		1	9	2	3	1		6	15
D2 'What I liked'	0	1		0	1	0	2	4		6	7
D3 Self-deprecation	14	2		0	16	3	0	19		22	38
D - Other	3	0		0	3	0	0	1		1	4
D - Total	25	3		1	29	5	5	25		35	64
ABCD Total	160	119		129	408	33	108	74		215	623
I 'Neutral' interviewer turns	43	59		66	168	12	36	37		85	253
R 'Neutral' respondent turns	57	56		51	164	13	39	27		79	243
GRAND TOTAL	260	234		248	740	58	183	138		379	1119

8. Sample subjective analyses (and transcript extract)

This section contains my impressionistic analyses, completed before coding (but after coding other interviews) and not thereafter edited, of what appear to be the two 'extreme cases' among the interviews analysed so far, followed by an extract from the transcript of one of the interviews, and summary comments on the others.

8.1 Interview 1

Respondent does sometimes set her own agenda, but the main content is persistent self-deprecation: the materials must be perfect, she must be stupid when she can't do something. Sometimes a hint of criticism of materials, but so mitigated by self-criticism as to be uninterpretable. Does not perceive interviewer as outsider - repeatedly says 'your' (notes etc.), ignores disclaimers.

Interviewer begins quite 'ethnographically', but quite soon moves (is forced?) into role of supportive teacher/'expert'. Last part is highly structured list of questions.

8.2 Interview 7

Worked well as ethnography. Interviewer made a convincing self-introduction: "I've already interviewed 3 people so I know a bit ... ". His invitation for general comments produced a series of R-initated topics (audio quality, self-discipline, importance of prediction etc.) with long R turns. R is confident enough not to blame herself for problems.

Works even better when, half-way through interview, they reach parts of the material that R did not fully understand. Extended mutual help: R contributes her knowledge of Italian, I his experience as a language teacher, and they solve problems and explore issues and perceptions together as equal partners.

A slightly shortened transcript extract is now given to illustrate the above.

Ron: *It says, "This is a ... " What's 'elenco', 'elenco'?*

Linda: *A list.*

Ron: *"... a list of ways of ..."*

Linda: *... to box. 'Attenuare' means to extenuate basically [...] that Andrea is using, is emphasising these phrases with words [...] and prepare a list of the way the ... these phrases have been used relating to the time, a film, a book that ... I just couldn't understand ... (indecipherable) ...*

Ron: *... (indec) ... "tua opinione" so you had to express your own opinion on something.*

Linda: *Yeah.*

Ron: *So maybe you just had to use these phrases in a sentence that you made up about something from your own experience, do you think?*

Linda: *… possibly, possibly. I … just couldn't get a sense of exactly what I was … what we were expected to do … em … and I didn't understand the example which was "non sono cattivi pero" and then what I presumed to be the response to that was "Lui non è stupido pero è … (indec) … non li capisci" I thought well he's not sort of, he's not being 'cattivi', he's being 'stupido' so why … why is that …*

Ron: *Excuse me … (indec) … I mean, I'm not sure what Giulia had in mind, but I think probably it's one of these substitution things. Now I teach English you see and I do something similar …*

Linda: *… and you use another word that means …*

Ron: *… (indec) … No, no. It could be something quite different 'cos you're using the grammatical structure …*

Linda: *… grammatical structure …*

Ron: *… so it's probably known … known … (indec) … so it's like saying 'he's not … but …' or 'she's not … but …'*

Linda: *'not … but …' …*

Ron: *… or 'they're not … but …' and you put in … (indec) …*

Linda: *Yeah. Even though … (indec) … doesn't understand certain things there*

Ron: *Yeah.*

Linda: *Yeah. (7, 12, 8 to 7, 13, 11)*

8.3 Other interviews

These ranged between the extremes above: interview 4 almost as 'bad' as interview 1, interviews 3 and 5 sometimes as 'good' as interview 7 but without the mutual exploration, and perhaps spoilt by some long interviewer turns, e.g.

> *"Sorry, before we go on - you didn't think it mattered that you were a wee bit wide of the mark? You don't think you would have perhaps understood more on the first hearing if you had kind of been spot on with your prediction?" (5, 3, 2)*

Interview 2 was rather a non-event as the respondent had only used a small part of the material; the other interviews are not yet analysed.

9. Conclusion

Until I analysed interview 7, I felt that nothing very clear had emerged on the meta-research issues. I had mentally reduced the factors affecting ethnographicity to three main ones - respondent's freedom to direct course of interview, respondent's perception of interviewer's areas of knowledge and ignorance, respondent's ideas about interviewer's allegiance and what interviewer wants to hear - but was still unclear about the relative importance of these and how they interact. I felt that we had been partially successful: having extensive experience of 'question schedule' interviews, I knew that the information yielded by these freer interviews was much richer, often more believable, and fulfilled the aim of insight into the life-world of the learners. But in other ways, especially in our attempt to disclaim expertise, we had obviously been clumsy and far from totally successful.

As soon as I read interview 7 I felt: this is the way to do it! It was exactly what ethnographers talk about, a 'lesson' given by Linda the interviewee. But it worked because the interviewer did *not* behave like a traditional ethnographer: he not only admitted, but frequently asserted, his expertise, but in such a way that Linda's expertise was also needed, so real communication took place. So, to my great surprise, I have (at least to my own satisfaction) a fairly clear answer to my question: no, applied linguists cannot do ethnography in their own field, but they can *at least sometimes* achieve a similar result by slightly different means.

I believe that similar approaches could have improved the already fairly satisfactory interviews (3 and 5) but still do not know how the others could have been "rescued". The sex (male) and age (42 and 54) of the interviewers probably contributed to the reaction of the (female) interviewees, and a same-sex interviewer, especially one more clearly distanced from 'authority', might have got better results.

It is impossible to provide any fixed algorithm or set of procedures for conducting ethnographic interviews: an element of 'playing by ear' is always involved. The use of systematic coding, as in the present study, should not be seen as an attempt to eliminate this element, but rather to emphasise the difficulty of this research technique and the need for interviewer self-awareness. Future researchers using ethnographic interviews should consider beginning with a pilot study and analysing pilot data from this perspective.

The coding system offered here is not claimed to be fully satisfactory: readers of an earlier draft of this article have suggested that certain categories may reflect individual interviewer style rather than ethnographicity or its opposite, and that the disregarding of paralanguage and of sequential patterns are limitations. It will be noticed that I have not, in the present article, appealed directly to evidence from the coding system, and indeed I found myself relying much more, or more immediately, on subjective analyses when formulating general conclusions. But the coding system has helped and (perhaps in revised form) will help in three ways: it guided my analysis to the point where I could take these impressionistic short-cuts; it has revealed detailed patterns of interest for a fuller report - for example, that in the 'successful' interview 7 leading questions were often followed immediately by respondent contradictions; and it will help me and, I hope, others to validate the provisional and perhaps premature conclusions and supply a more rounded picture.

Acknowledgement

This article would not have been possible without the work of Hélène Mulphin, Giulia Dawson, Stella Peyronel, and especially of my fellow interviewer Ron Howard. Any positive features are largely due to their efforts, any failings are my own.

References

Dawson G. and S. Peyronel. 1991. Italian Study Pack, Unit One. Unpublished teaching materials, available from IALS.

Howard R. 1992. '"Long interviews" on French/Italian study packs: a report'. Mimeo.

Malinowski B. 1922. Argonauts of the Western Pacific. London: Routledge.

Mulphin H. 1991. French Study Pack, Unit One. Unpublished teaching materials, available from IALS.

Spradley J. P. 1979. The Ethnographic Interview. New York: Holt, Rinehart and Winston.

INFLUENCE OF LANGUAGES OTHER THAN THE L1 ON A FOREIGN LANGUAGE: A CASE OF TRANSFER FROM L2 TO L3

Matutin Sikogukira (DAL)

Abstract

The phenomenon of transfer in language learning has mostly been investigated with reference to L1 and L2. This paper describes a case of transfer from L2 to L3, specifically the influence of French (L2) on the learning of English (L3). The study focuses on French-English lexical cognates and suggests that although the learners perceive French and English as closely related, they do not adopt a wholesale transfer strategy. Their assessment of the transferability of the cognates seems to depend on such factors as the category of cognates, the sense relations holding between cognates and other semantically related lexemes, and the learners' level of proficiency.

1. Introduction

One aspect of language transfer which, though not wholly neglected in recent literature, has nevertheless not yet captured the attention of most SLA researchers is that of the influence of languages other than the L1 on the target language. Most research on language transfer seems to assume that the natural route of transfer is from L1 to L2. Very little attention has been paid to the question of the extent to which languages other than the L1 influence the learning of an additional language. The way a learner with previous knowledge of another language acquires a new language will differ in some respects from that of monolingual learners in the same learning situation, with the same mother tongue and the same socio-psychological characteristics. Thomas's (1988) study, for instance, suggests that bilinguals learning a third language seem to have developed a sensitivity to language as a system which helps them perform better in those activities usually associated with formal language learning than monolinguals learning a foreign language for the first time. She argues that bilinguals who have formally acquired an L2 have developed a conscious awareness of language as a system that provides them with additional advantages over monolinguals and that their metalinguistic awareness may increase the potential advantage of knowing two languages when learning a third. Her findings also show that bilingual students learning a third language outperform monolingual students learning a second language.

There is wide agreement among SLA researchers whose work is centred on cross-linguistic influence that transfer (both positive and negative) is more likely to take place from a language which is related to the new foreign language being learned (see Corder 1979; James 1977; Kellerman 1987; Lightbown and Libben 1984; Nababan 1981; Ringbom 1978, 1986, 1987; Sweet 1964; Vildomec 1963). One researcher whose main interest is the notion of language similarity is Kellerman (1977, 1986, 1987). He argues for the psychotypology hypothesis, that is to say, the amount of transfer that a second language learner will attempt is determined in large measure by the learner's perception of the distance and the degree of relatedness and similarity between the source language and target language. According to him, learners may develop a notion of typological distance between the two languages by perceiving the source language as more or less distant from the target language. This perceived distance between the two languages together with the learner's fragmentary knowledge about a specific structural domain of the target language will allow the learner to make a prediction of the transferability of a source language feature.

If we assume that L2 influence on L3 is a reality, why is it, then, that L3 learners should be more ready to transfer from their L2 than from their L1? Corder (1979:33) points out that 'other languages known to the learner, however imperfectly, may, in the degree to which they resemble the target language, have a facilitating effect'. He goes on to argue that this assumption is supported by the general observation that 'the more languages one knows, the easier the acquisition of yet another appears to be' because in such a case 'the learner has a large number of "ready-made" hypotheses to test in processing the data of the new language'. He concludes that the magnitude of the task of learning an L2 which is related to one's L1 is much smaller than that of learning an unrelated language. He contends that where the mother tongue is formally similar to the target language the learner will pass more rapidly along the developmental continuum (or some parts of it), whereas in the case of unrelated (distant) languages the speed will be slower because of the differences along the whole continuum. Citing the example of Indonesian learners of English who transfer from their previously learned Dutch, in the areas of lexis and grammar, Nababan (1971) also claims that L2-L3 influence is common when the two languages are cognates.

Transfer of linguistic structures from the language which has greater resemblance to the target language among those known to a multilingual learner, rather than from his L1, has been referred to as 'the base language hypothesis' (Chandrasekhar 1978). He maintains that if a learner is multilingual, it is not always the mother tongue which interferes with the learning process but it may be another language. He contends that if the new language has greater resemblance to one of the languages known to the learner other than the mother tongue, it is from that language that transfer takes place and the possibilities of errors have to be determined by a contrastive analysis of this language and the new foreign language. This language from which transfer takes place, he calls 'the base language'. Tenjoh-Okwen's (1985) analysis of the interlanguages of francophone Cameroonian learners of English suggests that 44% of the deviant forms from the corpus analysed are attributable to French, 'the base language', and not to the learners' mother tongues.

The best-known work in the area of lexis is that carried out in Finland with bilingual Finnish-Swedish population. The name often associated with this research in this region is that of Ringbom (1978, 1983, 1986, 1987), whose analysis shows that the Finnish-Swedish learners of English as a foreign language significantly make more errors which are attributable to Swedish, than Finnish, irrespective of whether their L1 is Swedish or Finnish. He argues that Finnish learners of English rarely 'borrow' from Finnish; they prefer to 'borrow' from Swedish although they may resort to Finnish rather than Swedish when it is a question of a word's 'semantic field'. Ringbom (1978 : 96) stresses that it is sometimes claimed that when one speaks an L3 or an L4, influence from other foreign languages is much more apparent than L1 influence. He notes, however, that this view has so far been based on anecdotal evidence. In another study, Ringbom (1986:156) once more underlines that the extent to which languages other than the L1 influence the learning of an additional language has not yet been substantially investigated. This issue has so far been discussed in only a few scattered articles such as Ahukana et al. (1981), Chumbow (1981), LoCoco (1976), Ringbom (1978, 1986), Ulijn et al (1981) and unpublished theses (e.g. Bentahila 1975; Tenjoh-Okwen 1985; Wickstrom 1980) which are generally confined to exploring cross-linguistic influence in the area of lexis, usually between two related languages.

Other studies appear to refute Ringbom's view that influence from languages other than the L1 seems to be insignificant in the area of grammar and non-existent in phonology. In the area of syntax, for example, Khaldi (1982), in a study of acceptability judgment tasks on relative clauses and idioms by Algerian learners of English, compares learners from a bilingual setting with learners from an Arabic setting and finds that the bilingual learners transfer from their L2 (French) rather than from their L1 (Arabic) whenever they perceive the structure as language-neutral. He also notes that bilinguals perform better on the relative clause task because French rules are closer to English than Arabic ones. In a more or less similar study, Schachter et al. (1976) find that Arabic learners who are bilingual in French reject non-native-like relative clauses (in English) which resemble Arabic but not French, pointing to a case of positive transfer resulting from the application of L2 knowledge. White (1987), on the other hand, compares English-speaking learners of French and learners of French with other mother tongues but with previous knowledge of English and finds that the latter are more likely to accept preposition stranding in French. She argues that this might be due to transfer from English. In the area of phonology, Singh and Carroll (1979) show that their Indian informants are influenced by English rather than by their Indian L1s in their pronunciation of French. There is, however, a case of counter-evidence attested by Haggis (1973), who finds that Ghanaian Twi-speakers show far more evidence of Twi (L1) than English (L2) influence in their pronunciation of French. Perhaps most studies sugggest that L2-L3 influence is attested at all levels of language.

Although L2-L3 similarity is widely argued for in the literature as the cause for L2-L3 influence, it is not the only cause. L2-L3 influence seems to be an interplay of a number of factors. Bentahila (1975) and Rivers (1979) argue for 'recency' as a possible factor. This implies that whichever foreign language was learned last will interfere with the next-learned one. Meisel (1983) posits a 'storage and retrieval' factor and suggests that L2-L3 influence could result from the possibility that the way foreign languages are stored and processed in the brain may be different from the way first languages are stored and processed, irrespective of whether they are related or not. Vildomec (1963) underlines the learning style and setting (hence 'psychological similarity') by suggesting that if two languages are learned in a similar way by a similar method or in a similar situation, and if there is a similar emotional involvement with the milieu, they may influence each other. Finally, Singh and Carroll (1979) postulate a 'socio-cultural' reason by suggesting that L3 learners may identify more strongly with an L2 than with their L1, which could result in L2 influencing their learning of an additional foreign language. Although these factors may contribute to bringing about L2-L3 transfer, to different degrees, there is a wide agreement in the literature on cross-linguistic influence that L2-L3 transfer mostly occurs between similar or related languages. Some limited counter-evidence to this view has, nonetheless, been provided by some case studies (see Haggis 1973; LoCoco 1976).

2. The present study

The present research is a case study of the transferability of lexical properties from French as an L2 to English as an L3. It rests on the fundamental assumption that the transfer potential, pattern and process are determined not only by the degree of relatedness between the learner's L1 (or any other languages known to him) and the target language, but also the learner's perception of the distance between the source language(s) and the target language.

2. 1 The context

The language situation in Burundi can be regarded as particularly favourable for investigating how the transfer phenomenon is influenced by the above mentioned two factors. As far as learning English is concerned, all students' command of English is very much a knowledge of a foreign language rather than a second language since all of them are bilingual, having Kirundi as their L1 and French as their L2. As part of my teaching experience in the Department of English Language and Literature at the University of Burundi, I have observed that Burundian students of English make a comparatively large number of semantic approximations due to transfer of the semantic structure of the L2 (French). This seems to indicate that the frequency of such lexical errors is much influenced by the relatedness of French (L2) to English (L3). There is little doubt that reliance on word form and morphemic similarities between two related languages can lead to errors, although we can only have clues to the underlying process when learners go wrong. The underlying assumption is that, by virtue of the genetic relatedness and, hence, the formal and semantic similarities between French and English lexical items, Burundian students of English transfer more readily lexical properties from French to English rather than from Kirundi to English. French here functions as the 'base language'.

2.2 The subjects

The subjects involved in this study were 126 students of the Department of English Language and Literature at the University of Burundi (50 from first year, 28 from second year, 25 from third year and 23 from fourth year, with an average age of 22, 23, 24, and 25 years respectively). There are three main reasons for choosing this particular population as subjects of the experiment.

First, they all share the same linguistic, cultural and educational background in that they have the same mother tongue (Kirundi), have been taught in French and followed the same national curriculum throughout primary and secondary education, and were raised in an exceptionally monocultural speech community. It is hoped that this homogeneity factor will increase the degree of generalisability of the results. Second, and perhaps more importantly, all the subjects have experienced the same training in English prior to their entry to the Department of English Language and Literature. They have taken English for six years in secondary education following the same national curriculum and are now attending a four year course in the above mentioned Department where the sole medium of instruction is English. Their admission to the Department is dependent on their performance in English in a national test administered at completion of secondary education whose aim is to determine the potentially best candidates for each academic discipline. Thus it is understood that the majority of them must have achieved the best performance in English nationwide and may be regarded as the best models of the English language in the entire country.

Additionally, they have been taught English by staff who are nearly 100% locally trained nationals of Burundi who have exactly the same linguistic, cultural and educational background as the students themselves. In the Department of English Language and Literature, they continue to be taught by national academic staff, except two or three foreign staff members. The point that is being emphasized is that we are dealing with francophone learners of English who have been trained by francophone teachers, who are non-native speakers of French, in a predominantly Kirundi environment with the result of students' interlanguages being to some extent the product of their teachers' own interlanguages. This further factor may increase the chances of occurrence of 'Frenchisms' in students' performance in English.

Third, the subjects belong to four different years of study (first, second, third and fourth years), leading to the award of the degree of 'Licence' (equivalent to B.A) in English Language and Literature. Therefore they have different levels of proficiency in English. At the same time it seems that the four different levels of study could correspond to different levels of students' linguistic and metalinguistic awareness according to the structure of the Department curriculum, since some courses which are intended to enhance students' linguistic and metalinguistic awareness are postponed till students have been introduced to some other course entries. For instance, general linguistics is taught in first year, descriptive grammar and practical phonetics in second year, syntax, semantics, evolution of the English language and phonetics and phonology in third year, psycholinguistics, sociolinguistics and advanced topics in linguistics in fourth year. It is therefore hypothesised that the subjects' responses will vary according to the level of proficiency.

2.3 The lexical category

The lexicon is such a huge and multidimensional network that to tackle all of it within the confines of the present paper would be neither desirable nor feasible. Thus it is necessary to select and delimit a manageable lexical category that can satisfactorily attest lexical transfer from French to English. The lexical category of French-English cognates has been selected for this purpose. Not only does this category cover a large common lexicon between French and English but it is also potentially exceptionally rich for investigating the transferability of lexical properties from French to English. Moreover, it is well known that in language learning situations involving closely related languages, cognates always baffle language teachers and learners because language teaching coursebooks and textbooks generally fail to propose an appropriate methodology for the teaching and learning of cognates. As a matter of fact, cognates occur in many guises which may not always be easy for learners to identify. However, although cognates constitute nasty pitfalls in language learning, they are also a useful asset for rapid vocabulary acquisition and development of lexical knowledge.

Nevertheless, even the category of French-English cognates remains too broad a topic to be dealt with at one time. Since the major problem inherent in the use of cognates lies essentially in the assessment of their semantic overlap or semantic difference between language x and language y, it seemed appropriate that this study should concentrate on the semantics of French-English lexical cognates, and not deal with their morphology. Accordingly, two types of categories have been selected.

The first category includes French-English cognates whose meanings are the same or similar in French and English and which are in a relation of synonymy with non-cognate English lexemes (e.g. *commence, begin, start; espionage, spying*) or hyponymy (e.g. *assassinate, murder, kill; gluttony, gourmandise, greed*). By regarding cognates as cross-linguistic synonyms despite their usage differences, we accept that synonyms serve two important and complementary functions in everyday communication. First, they add flexibility to the language by enabling its users to express the same meaning by different means. Second, they add variety and expressiveness to the language by enabling its users to exercise stylistic choices in conveying the same message (see Chi-wei 1983). On the other hand, hyponymy as a semantic relation of inclusion whereby the meaning of a more specific lexeme is included in that of another more general lexeme allows the possibility of avoiding repetitions, defining or describing concepts through hyponymous substitutions. It is often argued by semanticists (e.g. Lyons 1981) that language users are likely to know the superordinate terms and their full meanings but do not necessarily know the full meanings of their corresponding hyponyms although they perceive a certain semantic link between them. In this study, it will be shown that synonymy and hyponymy are important sense relations which underly the selection and use of French-English cognates by Burundian university students of English.

The second category includes French-English cognates whose meanings differ in the two languages (e.g. *venue, siege, tutor*). This is the classic category of lexemes that most theoreticians, especially those whose work has pedagogical aims, usually have in mind when they talk of false cognates. In this study, it will be shown that this is by far the most difficult and treacherous class of cognates in the sense that learners tend to anticipate a semantic similarity where they see a formal one.

In order to minimise extraneous factors that can further obscure the phenomenon of cognateness, it is important that we limit our study to simple cognates and leave out complex cognates such as derivatives and compounds as far as possible. The latter may indeed involve different kinds of knowledge and their acquisition may therefore appear to be more complex than that of simple cognates. Although a few derived cognates which are commonly acknowledged as classic examples of French-English false cognates such as *actually* and *eventually* will be included in our data, word-formation and derivational morphology is not the concern of this study.

2.4 The hypotheses

French and English share a large common lexicon mainly as a result of the contacts the two languages have had in the course of time. Each of the two languages has borrowed words from the other, but rarely have these words kept the original meaning in the borrowing language. False cognates generally result historically from semantic shift in the sense that once a lexical item is present in two languages, its meaning can alter or diverge in various ways : it can be restricted (e.g. *commence* is used in formal contexts in English but not in French), it can be added to (e.g. *venue*, which denotes the action of coming in French, denotes the place of the action in English), etc. Language learners often have little or no training in historical linguistics and usually expect a semantic similarity where they see a formal one between pairs of cognates in two languages. Even when such a similarity does exist, learners may mistrust it and adopt an avoidance strategy by simply not venturing to use the cognates in question. The present study aims to investigate some generalisable ways in which Burundian university students of

English handle French-English cognates, that is, the factors underlying their decisions to transfer or not to transfer their knowledge of the cognates in French into English. Accordingly, the following hypotheses correspond to my predictions about the subjects' use of the above mentioned categories of French-English cognates.

Burundian university students of English will

1. show a tendency to use non-cognate English lexemes which are in a relation of either synonymy or hyponymy with French-English cognates, rather than the latter,
2. show a tendency to transfer French-English cognates whose meanings differ between French and English,
3. show a variation of their behaviour in 1 and 2 according to their knowledge of English : in both cases the tendency decreases with the increase in their level of proficiency.

2.5 The experiments

2. 5. 1 Experiment 1 : sentence completion task (see appendices A and B)

The subjects were presented with sentences in which a word was missing, and were required to supply the missing word. Although the sentences provided as much information as possible and used contexts which were familiar to the subjects to facilitate their guessing, there were risks of subjects' misunderstanding or misinterpreting the contextual and intended meaning. As a possible way to control these variables, the test was administered in two slightly different versions. The first version required the subjects to find the omitted word by relying exclusively on the information provided by the context of the sentence (see Appendix A). In the second version, the subjects were presented with the same sentences, this time with French translations for the omitted words to constrain them to make their choices within limited lexico-semantic boundaries (see Appendix B). The translations had a specific purpose because they were French-English cognates, most of the English equivalents of which were the correct words to use, and the experiment aimed at finding out whether the subjects would use the cognates or what other kinds of words they would tend to use instead.

Results

The table below presents the distribution of the words which were provided by the subjects as their answers and the percentages of the subjects who gave the words in each class. The version without French prompts will be referred to as V1 and the version with French prompts will be referred to as V2. French-English cognates are marked with a + in the table.

		year 1 (50 subjects)		year 2 (28 subjects)		year 3 (25 subjects)		year 4 (23 subjects)	
		V1	V2	V1	V2	V1	V2	V1	V2
1	misunderstanding	60	14	50	7.14	36	8	21.73	4.34
	disagreement	16	0	0	0	0	0	0	0
	break	14	64	25	42.85	36	40	34.78	34.78
	split	0	6	3.57	14.28	0	8	0	8.69
	rupture+	4	6	14.28	35.71	20	36	39.28	47.82
	others(cut,clash)	6	10	7.14	0	8	8	4.34	4.34
2	spying	80	78	71.42	71.42	56	48	43.47	43.47
	espionage+	10	16	21.42	28.57	40	44	56.52	56.52
	others (lying, betrayal)	10	6	7.14	0	4	8	0	0
3	greed-iness	76	82	46.42	50	32	32	30.43	30.43
	behaviour	2	0	10.71	0	4	0	0	0
	over-eating	10	10	3.57	0	16	8	4.34	0
	selfishness	6	4	3.57	7.14	0	4	0	0
	gluttony+	0	2	14.28	14.28	24	28	26.08	26.08
	gourmandise+	2	2	17.85	25	20	28	34.78	43.47
	others (queerness)	4	0	3.57	3.57	4	0	4.34	0
4	left	46	12	32.14	21.42	28	24	26.08	13.04
	stopped	36	66	32.14	42.85	28	32	21.73	21.73
	interrupted+	10	20	28.57	35.71	40	44	52.17	56.52
	others (gave up)	8	2	7.14	0	4	0	0	8.69
5	loan	80	82	60.71	64.28	60	60	56.52	56.52
	credit+	16	18	32.14	32.14	36	40	43.47	43.47
	others (lending)	4	0	7.14	3.57	4	0	0	0
6	way	74	70	67.85	60.71	52	52	43.47	43.47
	road	12	12	10.71	10.71	8	8	4.34	4.34
	street	4	8	0	3.57	0	0	0	0
	route+	10	10	21.42	21.42	40	40	52.17	52.17

7	beginning	92	96	78.57	89.28	72	80	65.21	73.91
	opening	4	2	7.14	0	8	4	13.04	8.69
	start	4	2	7.14	0	8	4	4.34	0
	commencement+	0	0	7.14	10.71	12	12	17.39	17.39
8	killed	62	66	50	42.85	32	32	26.08	26.08
	murdered	16	16	17.85	28.57	28	28	26.08	26.08
	shot	12	8	7.14	0	8	4	4.34	0
	assassinated+	10	10	25	28.57	32	36	43.48	47.84
9	freed	62	62	39.28	46.42	24	28	34.78	34.78
	released	8	10	21.42	14.28	20	12	0	0
	liberated+	24	22	35.71	35.71	56	60	65.21	65.21
	others (blessed)	6	6	3.57	3.57	0	0	0	0
10	experiments	28	34	53.57	60.71	72	72	86.95	91.13
	experiences+	72	66	46.42	39.28	28	28	13.04	8.69
11	deposited+	0	6	10.71	25	32	32	39.13	39.13
	put	68	72	42.28	46.42	28	28	17.39	30.43
	kept	10	12	7.14	3.57	8	12	8.69	4.34
	saved+	16	10	17.85	17.85	24	24	30.43	26.08
	others (sent)	6	0	10.71	3.57	8	4	4.34	0
12	tiredness	64	66	53.57	53.57	32	32	26.08	26.08
	fatigue+	30	34	42.85	46.42	68	68	73.91	73.91
	others (thirst)	6	0	3.57	0	0	0	0	0
13	deposit	6	8	17.85	21.42	32	32	43.47	47.82
	caution+	84	82	60.71	71.42	48	56	52.17	47.82
	warranty	0	0	10.71	7.14	16	12	4.34	4.34
	others (sureness)	10	10	10.71	0	4	0	0	0
14	physicists	6	8	21.42	25	40	40	56.52	56.52
	physists*	0	0	3.57	7.14	16	16	8.69	13.04
	physicians+	82	90	64.28	64.28	40	44	17.39	21.73
	scientists	10	2	3.57	3.57	4	0	17.39	8.69
	others (scholars)	2	0	7.14	0	0	0	0	0
15	deranged+	0	6	10.71	10.71	20	20	26.08	34.78
	disturbed	70	72	50	39.28	32	32	17.39	13.04
	damaged	10	0	7.14	14.28	12	12	17.39	13.04
	troubled+	20	16	32.14	32.14	36	36	34.78	39.13
	others (unsettled)	0	6	0	3.57	0	0	4.34	0
16	involved	72	74	50	50	40	40	26.08	26.08
	implicated+	14	14	25	28.7	48	48	73.91	73.91
	included+	6	6	25	21.42	12	12	0	0
	others (showed)	8	6	0	0	0	0	0	0

17	goods	76	80	50	53.57	44	40	34.78	34.78
	things	6	0	7.14	7.14	0	0	0	0
	items	4	0	10.71	3.57	0	0	0	0
	products+	10	10	17.85	21.42	20	20	21.73	21.73
	merchandises+	4	10	14.28	14.28	36	40	43.47	43.47
18	give back	64	68	57.14	50	44	40	30.43	30.43
	pay back	14	14	10.71	17.85	20	24	13.04	13.04
	return+	10	10	14.28	21.42	20	16	17.39	17.39
	reimburse+	4	4	7.14	10.71	16	20	39.13	39.13
	others (bring)	8	4	10.71	0	0	0	0	0
19	deeply	84	94	71.42	71.42	56	60	30.43	30.43
	sound	8	2	0	0	0	0	0	0
	profoundly+	8	4	28.57	28.57	44	40	69.56	69.56
20	car	96	94	82.14	82.14	68	72	43.47	43.47
	vehicle+	4	6	17.85	17.85	32	28	56.52	56.52
21	carelessness	80	82	50	53.57	48	48	34.78	30.43
	negligence+	8	10	35.71	35.71	40	40	60.86	60.86
	neglect+	4	2	14.28	10.71	12	12	4.34	8.69
	others (betrayal, wrong doing)	8	6	0	0	0	0	0	0
22	begins	80	88	39.28	42.85	40	40	47.82	47.82
	starts	20	12	46.42	42.85	44	44	21.73	21.73
	commences+	0	0	14.28	14.28	16	16	30.43	30.43
23	left	70	72	60.71	57.14	48	40	21.73	26.08
	gave up	6	12	3.57	10.71	8	12	4.34	8.69
	abandoned+	16	14	32.13	32.13	44	44	69.56	65.21
	others (forsook)	8	2	3.57	0	0	4	4.34	0
24	team	88	88	82.14	82.14	80	76	73.91	73.91
	club	6	6	3.57	3.57	4	4	4.34	4.34
	formation+	6	6	14.28	14.28	16	20	21.73	21.73
25	end	70	88	53.57	75	48	64	47.82	47.82
	begin/start	22	0	14.28	0	12	0	0	0
	finish+	8	12	21.42	14.28	24	20	30.43	30.43
	terminate+	0	0	10.71	10.71	16	16	21.73	21.73
26	surrendered+	14	14	39.28	32.14	40	36	43.47	43.47
	withdrew	48	54	25	25	20	20	17.28	13.04
	gave up	28	22	0	10.71	8	4	4.34	0
	capitulated+	8	8	25	28.57	28	36	34.78	39.13
	others (lost)	2	2	10.71	3.57	4	4	0	4.34
27	introduce	26	20	53.57	57.14	76	76	82.60	86.95
	show	8	2	0	0	0	0	0	0
	present+	58	76	35.71	42.85	20	24	17.39	13.04
	others (name)	8	2	10.71	0	4	0	0	0

28 take	42	44	32.14	32.14	32	36	21.73	21.73
use	20	20	17.85	17.85	4	4	8.69	8.69
have	18	16	10.71	7.14	16	8	8.69	8.69
occupy+	16	18	32.14	42.85	48	52	56.52	60.86
others (fill)	4	2	7.14	0	0	0	4.34	0
29 bring	34	32	7.14	0	12	4	8.69	0
give	42	48	25	14.28	20	28	13.04	26.08
hand	8	6	3.57	7.14	0	0	0	0
pass+	16	14	64.28	78.57	68	68	78.26	73.91
30 explain	92	100	85.71	85.71	76	80	78.26	78.26
explicate+	0	0	10.71	14.28	20	20	21.73	21.73
others (grasp)	8	0	3.57	0	4	0	0	0

2. 5. 2 Experiment 2 : lexico-semantic acceptability judgment task (see Appendix C)

The subjects were presented with complete sentences containing a cognate word. The cognate, which was underlined, was appropriately used in some cases and in some other cases it was not. In other words, the experiment included sentences where the cognates were used according to their English meaning and others where they were used incorrectly, that is, according to their French meaning. The subjects' task consisted of giving their acceptability judgment for each case, that is, whether they accepted the use of the cognate as appropriate or did not. Since there were cases where the subjects might have felt uncertain about the acceptability of the use of the cognates, a yes/no or acceptable/unacceptable answer would have failed to show this indeterminacy. Therefore they were given a scale of five points along which they could rank their judgments. Point 5 meant completely acceptable, point 1 meant completely unacceptable and 4, 3, and 2 were intermediate points. Twenty-eight items used in this experiment relate to the first hypothesis, that is, use of non-cognate English words rather than French-English cognates. The other twenty-two relate to the use of cognates whose meanings differ in French and in English.

Results

The table below presents the average means of the answers given by the subjects in each class. The figures correspond to the subjects' tendency to accept (if close to 5) or not to accept (if close to 1) the use of each item.

	year 1 (50 subjects)	year 2 (28 subjects)	year 3 (25 subjects)	year 4 (23 subjects)
1 veterinary*	4.28	3.143	3.07	3.09
2 demanded	2.28	2.428	2.96	3.391
3 ignore	2.12	2.464	2.6	2.913
4 remarked	1.8	2.428	2.92	3.043
5 attained	2.14	2.464	2.76	3.174
6 attended	1.7	2.25	2.44	2.826
7 termination	1.82	2.214	2.6	3.00
8 devastated	1.82	2.25	2.64	3.261
9 succeeded	1.82	2.25	2.6	3.304

10 cautioned	1.82	2.25	2.56	2.913
11 administrates	2.38	2.428	2.76	3.304
12 saluted	2.28	2.464	2.92	3.478
13 commended	1.8	1.964	2.64	2.695
14 fatigued	2.06	2.607	2.64	3.130
15 venue	1.8	2.214	2.56	2.869
16 nominated	1.9	2.25	2.56	2.826
17 sympathetic	1.9	2.214	2.48	2.695
18 reprimand	2.28	2.392	2.96	3.434
19 ameliorate	2.18	2.464	2.84	3.478
20 inexcusable	1.86	2.357	2.68	3.130
21 recompense	1.86	2.464	2.8	3.434
22 entourage	1.94	2.392	2.76	3.434
23 theatre	1.82	2.285	2.44	2.695
24 grave	1.86	2.285	2.55	3.043
25 interests*	3.48	3.25	2.92	2.347
26 liberty	2.46	2.357	3.28	3.391
27 siege*	3.50	3.142	3.00	2.609
28 massive	2.38	2.392	2.56	3.217
29 necessitated	2.00	2.464	2.88	3.478
30 aid	1.9	2.214	2.6	3.347
31 estimate*	4.1	3.785	3.28	3.00
32 agenda*	4.76	4.179	3.00	2.695
33 depose	1.88	2.214	2.88	3.434
34 promenaded	1.6	1.785	2.56	3.086
35 comprehend 1	2.04	2.357	2.8	3.130
36 persuaded*	4.26	4.142	3.36	3.217
37 revenue	2.18	2.464	2.56	3.304
38 authoritative*	4.76	4.142	4.12	3.347
39 comprehend 2	2.08	1.928	2.6	2.695
40 alleges*	4.28	3.857	3.68	3.260
41 assassin	1.8	2.321	2.56	3.130
42 pardoned	1.9	2.285	2.52	3.086
43 function	2.00	2.321	2.4	2.869
44 actuality*	4.00	3.25	2.92	2.826
45 menace	1.72	2.464	2.96	3.391
46 guardian	2.00	2.392	2.56	3.478
47 concussion*	3.98	3.57	3.36	3.260
48 chanting	1.9	2.464	2.8	3.478
49 administered	1.80	2.214	2.428	2.695
50 formidable	1.6	1.928	2.56	3.086

Note: The words marked with a * are unacceptable in the contexts they are supplied in the experiment.

2. 5. 3 Discussion of the results

In the first experiment, all the items except number 10, 13, 14, and 27 relate to the hypotheses that the subjects will show a tendency to use non-cognate English lexemes which are in a relation of either synonymy or hyponymy with French-English cognates rather than the latter (hypothesis 1) and that this tendency will decrease with the increase in the subjects' level of proficiency (hypothesis 3). The four remaining items (number 10, 13, 14, and 27) correspond to the hypotheses that the subjects will show a tendency to transfer French-English cognates whose meanings differ between French and English (hypothesis 2) and that this tendency will decrease with the increase in the subjects' level of proficiency (hypothesis 3).

Regarding the first and third hypotheses, the evidence from the results rests on the comparison of the percentage of the subjects who used French-English cognates with the percentage of the subjects who used non-cognate English words and the comparison of the subjects' answers according to their level of proficiency. The first step is to identify which items among the answers given by the subjects are French-English cognates and which ones are non-cognate English words. We shall regard as French-English cognates all the items whose form is entirely or partially similar in French and English. These are distinguished in the table by a +. It should be noted, however, that there is no systematic way of measuring formal similarity, although common roots and affixes are reliable indicators of formal similarity between cognate pairs. On the other hand, we regard as non-cognate English words all the items which have no counterparts in French which are entirely or partially similar to them in form.

Overall, two important observations arise from the results in both version one (V1) and version two (V2) :

The subjects' answers are mostly non-cognate English words which are in a relation of either synonymy or hyponymy with the French-English cognates in question. However, the percentage of the subjects who used non-cognate English words decreases from left to right, i.e. from first year to fourth year, while the percentage of the subjects who used French-English cognates rises from right to left, i.e. from fourth year to first year. For example, in sentence number one, 60% and 64% of first year subjects used *misunderstanding* and *break* respectively in V1 and V2, while only 50% and 42.85% of second year subjects, 36% and 40% of third year subjects, and 21.73% and 34.78% of fourth year subjects did so. In the same sentence, 39.28% and 47.82% of fourth year subjects used *rupture* respectively in V1 and V2 whereas only 20% and 36% of third year subjects, 14.28% and 35.71% of second year subjects, and 4% and 6% of first year subjects used it. In sentence two, 80% and 78% of first year subjects used *spying* respectively in V1 and V2 where 71.42% of second year subjects , 56% and 48% of third year subjects, and 43.47% of fourth year students used it. Conversely, 56.52% of fourth year subjects used *espionage* rather than *spying* respectively in V1 and V2 where 40% and 44% of third year subjects, 21.42% and 28.58% of second year subjects, and 10% and 16% of first year subjects did so. The same kinds of proportions are observed in all the twenty six sentences. Therefore the results of the experiment support hypotheses one and three.

Regarding the four other items which relate to cognates whose meanings differ in French and in English, the subjects tend to transfer their French knowledge of the cognates into English but this tendency decreases with the increase in the subjects' level of proficiency. In sentence number ten, for instance, 72% and 66% of first year subjects used *experiences* where 46.42% and 39.28% of second year subjects, 28% of third year subjects, and 13.04% and 8.69% of fourth year subjects did so respectively in V1 and V2, whereas all the remaining subjects used *experiments*.The same observation applies to *caution* in sentence 13, *physicians* in sentence 14, and *present* in sentence 27. These results support hypotheses two and three.

The effect of French prompts in V2

As had been anticipated, in some cases, a number of subjects misunderstood or misinterpreted the sentences, this resulting in the subjects' failing to use the word which was expected, particularly in the version without French prompts. The subjects' answers in the version with French prompts did not significantly alter the subjects' tendency to use non-cognate English words rather than French-English cognates or to transfer cognates whose meanings differ in French and in English, although the figures in V1 and V2 are different for some items. The French prompts simply made it easier for the subjects to use the words expected but they also seem to have increased the subjects' likelihood of using French-English cognates rather than non-cognate English words. The words which changed the intended meaning of the sentences belong to the category of 'others' in the results table. In any case, they are so few as to bear no significance for the results of the experiment.

The second experiment comprises two categories of items :

(a) French-English cognates whose meanings are the same or similar in French and English : *demand, remark, attain, termination, devastated, administrate, salute, fatigued, reprimand, ameliorate, inexcusable, recompense, entourage, grave, liberty, massive, necessitate, aid, depose, promenade, comprehend 1, revenue, comprehend 2, assassin, pardon, menace, chant, and formidable.*

(b) French-English cognates whose meanings differ in French and in English : *veterinary, ignore, attend, succeed, caution, commend, venue, nominate, sympathetic, theatre, function, interest, siege, estimate, agenda, persuaded, authoritative, allege, actuality, guardian, concussion, and administer.*

The results of the experiment indicate that, for the first category of cognates, the mean of the subjects' rating of their acceptability rises from left to right, i.e. from first year to fourth year. They also indicate that all first and second year subjects rated their acceptability below 2.5 (except for *fatigued*), all third year subjects between 2.5 and 3 (except for *liberty*), and all fourth year subjects between 3 and 3.5 (except for *comprehend* 2). Yet all the items were acceptably used according to five native speakers (all applied linguists) whom I asked to give their acceptability judgments of the items to confirm my own intuitions, prior to running the experiment. Interpreted in the light of the stated hypotheses, the results suggest that the subjects tend to reject or avoid using French-English cognates whose meanings are the same or similar and that this tendency decreases with the increase in the students' level of proficiency.

Under the category of cognates whose meanings differ in French and in English, we have included false friends (e.g. *venue, sympathetic*), polysemous words (e.g. (*succeed, theatre*), and synforms (same lexical forms) or confusable pairs (e.g. *authoritative/authoritarian, estimate/esteem*). Laufer (1988, 1989) refers to this category as 'deceptively transparent words'. The results of the experiment show that those which were appropriately used, or to put it differently, those which were used in agreement with their English meaning, were poorly rated by the subjects from all the four classes (e.g. *ignore* : 2.12, 2.464, 2.6, and 2.913; *attended* : 1.7, 2.25, 2.44, and 2.826; *cautioned* : 1.82, 2.25, 2.56, and 2.913; *venue* : 1.8, 2.214, 2.56, and 2.869; etc.), whereas the ones which were inappropriately used, or to put it differently, those which were used compatibly with their French meaning but incompatibly with their English meaning, were highly rated by the subjects from all the four classes (e.g. *interest* : 3.48, 3.25, 2.92, and 2.347; *siege* : 3.50, 3.142; 3, and 2.260; *agenda* : 4.76, 4.179, 3, 2.695; *persuaded* : 4.26, 4.142, 3.36, and 3.217; *alleges* : 4.24, 3.857, 3.68, and 3.260). The subjects' acceptability judgments seem to have depended on whether or not the meaning of the words in the contexts they were used in coincided with or differed from the one they assign to the words in French. With polysemous cognates, for instance, they seem to have assumed that succeed means only 'manage to', that theatre has to do with only 'plays', that administer has only to do with 'manage' or 'run', and they substituted interest for 'profit' as they belong to the same semantic field although they do not mean the same thing.

Among confusable pairs, *veterinary* was taken for 'veterinarian' because they both have the same French equivalent 'vétérinaire' and was highly rated by all the groups (4.28, 3.143, 3.07, and 3.09), *estimate* (3.82, 3.785, 3.25, and 3) was confused with 'esteem' because they share the same French equivalent 'estimer', and *authoritative* was confused with 'authoritarian' because they are both related to the French word 'autorité' (authority) and was highly rated by the subjects from all the four classes (4.76, 4.142, 4.12, and 3.347). However, whether the subjects tend to accept or reject the use of the cognates, the results of the experiment show that this tendency decreases with the increase in the subjects' level of proficiency. Therefore the results support hypotheses two and three.

2. 5. 4 Interpretation of the Results

There are at least three possible reasons for the learners' avoiding using cognate words whose meanings are the same or similar in French and English. They may be doing so because they feel that the non-cognate English words semantically represent the concepts they stand for more precisely than cognate words do; or because they are deliberately adopting an avoidance or non-transfer strategy, especially when they lack confidence about the acceptability and appropriateness of French-English cognates; or else because they simply do not know the correct usage of the cognates in English (ignorance).

On the other hand, to explain why French-English cognates whose meanings differ between French and English seem to present more difficulties to the learners, one has to look at the hierarchy of difficulty involved in learning word meanings. In this particular case, the difficulty can be described as learning new meanings for known words, on the one hand, and learning new formal representations for known words, on the other hand. In other words, the subjects already know the words and their meanings in French but have to realise that these words denote different concepts in English; again, even if the concepts that the words denote in English are already known to the learners, they have the task of learning new labels for those concepts. And although the learners already know these labels in French, they also have the task of learning the differences between the labels (orthographic, morphemic, grammatical, etc.) in French and in English. Therefore such words are a potential source of difficulty. This difficulty is twofold because it involves expanding the meanings of words that the students already know in the source language (in this case, French) by acquiring additional meanings that the words have in the target language (in this case, English) and learning to differentiate between two formal representations (the French and the English) of the same underlying word. The question which remains unanswered is, however, how the forms and meanings of such cognates are stored and coexist in the mental lexicon and what processes are involved in accessing and retrieving them while performing in either language. From a semantic point of view, cognates whose meanings differ between language x and language y can be described as 'cross-linguistic polysemous items', with the implication that the difficulty involved in learning and using 'intra-linguistic polysemous items' also applies to cross-linguistic polysemous items.

Since the subjects who took part in the experiments belonged to four different groups (first, second, third, and fourth years), it was useful to observe whether there was any significant variation in their performance behaviour. It was predicted that the subjects' tendencies to use synonymous or hyponymous non-cognates rather than French-English cognates and to transfer French-English cognates whose meanings differ in French and English would both decrease with the increase in the subjects' level of proficiency. This appears to be borne out by the data. The reason for this variation in the subjects' behaviour is twofold. On the one hand, learners' performance in the target language is naturally expected to improve as their level of proficiency increases. On the other hand, we can explain the variation in the specific area of lexis in terms of the organisation of the bilingual lexicon and the principles of word recognition and retrieval which continue to undergo some restructuring along the target language developmental route in such a way that bilingual individuals with different levels of proficiency in the target language presumably have their mental lexicon organised differently and use different word recognition and retrieval models.

In terms of language learning theory, the above results imply that the level of proficiency is an important factor which influences the learners' performance in the target language. On the one hand, it is often argued that beginning or less advanced learners are biassed towards the source language and are attracted to formal similarity but are less successful in working out semantic similarity in cognate pairs, whereas advanced learners make target language-based associations. In other words, advanced learners make semantic associations within the target language. This may also imply that as learners progress and their confidence in the target language grows, they gradually move away from the source language and possibly start 'thinking' in the target language. On the other hand, the results of this study show that this is not always the case. For example, as far as French-English cognates are concerned, it appears that the proficiency factor interacts with the category of cognates being considered. In terms of communication efficiency, the subjects' use of synonymous or hyponymous alternatives to the cognates may result in lack of communicative precision as a consequence of semantic approximations. For instance, a hyponym and its superordinate counterpart do not cover the same area of meaning and would not be interchanged in most contexts without resulting in semantic imprecision and communicative inaccuracy.

3. Further Implications for SLA

From the above observations, it appears that the subjects are suspicious of some categories of cognates but not of others. First, they tend to avoid using French-English cognates which have synonymous or hyponymous non-cognate English alternatives and to use the latter instead. Second, they tend to transfer more readily French-English cognates whose meanings differ in French and in English. Does this suggest that the first category is perceived as less transferable than the second category by the learners? Does it suggest that the learners use different recognition and retrieval strategies for different categories of cognates? Is it the case that different categories of cognates may be arranged in different sub-components of the mental lexicon? Does it suggest that bilingual individuals have two separate mental lexicons and that some categories of cognates are incorporated in one of the two lexicons whereas some other categories are incorporated in the other lexicon? And if bilingual individuals have only one common lexicon for both languages, what are the underlying factors which determine some cognates being more transferable than others? Does it also imply that the notions of psychotypology and language distance interplay with other factors such as the linguistic (in this case, lexical) categories being considered, the semantic relations holding between lexical items, and the learners' level of proficiency? Is it therefore insufficient to assume that the learners' perception of the distance between the source language and the target language will automatically boost or depress the likelihood of transferability? And finally, does it imply that different strategies need be used to teach different categories of cognates? It is these questions that make French-English lexical cognates an interesting and important area of investigation, and it is an awareness of the relevance of these questions that has motivated the present study.

4. Conclusion

It appears from this study that the strong belief among SLA researchers working on lexical transfer (e.g. Haastrup,1989; Ringbom, 1987) that 'we do well in letting learners understand that lexical transfer is overwhelmingly positive ... when the L1 and L2 in question are related ... ' is valid only with regard to some lexical categories. The transferability of French-English cognates largely varies in accordance with the lexical categories, the semantic relations holding between cognate pairs/sets and the learners' level of proficiency. This study provides further evidence for L2 influence on L3, but I believe that more studies should be carried out to confirm other cases of L2-L3, or even L3-L4, influence before this research area can gain more ground.

References

Ahukana J., N. Lund and J.R. Gentile. 1981. 'Inter- and intra-lingual interference effects in learning a third language'. Modern Language Journal 65 : 281-87.

Andersen R.W. 1984. Second Languages : A Crosslinguistic Perspective. Rowley, Mass. : Newbury House.

Bentahila A. 1975. 'The Influence of L2 on the Learning of L3'. M.A Dissertation. University College of North Wales, Bangor.

Chandrasekhar A. 1978. 'Base language'. IRAL 16/1 : 62-5.

Chi-wei T. 1983. Towards an Explication and Description of Synonymy in English'. Ph.D Thesis. University of Edinburgh.

Chumbow B.S. 1981. 'The mother tongue hypothesis in a multicultural setting' in J.G. Savard and L. Laforge (eds.) 1981 : 42-55.

Corder S.P. 1979. 'Language distance and the magnitude of the learning task'. Studies in Second Language Acquisition 2/1 : 27-36.

Haastrup K. 1989. 'The learner as word processor'. AILA Review 6 : 34-46.

Haggis B.M. 1973. 'Un cas de trilinguisme'. La Linguistique 9/ 2 : 37-50.

James J. 1977. 'Language transfer reconsidered'. Interlanguage Studies Bulletin 2/3 : 7-21.

Kellerman E. 1977. 'Towards a characterisation of the strategy of transfer in second language learning'. Interlanguage Studies Bulletin 2(1) : 58-146.

Kellerman E. 1986. 'An eye for an eye : cross-linguistic constraints on the development of the L2 lexicon' in E. Kellerman and M.Sharwood Smith (eds.) 1986 : 35-48.

Kellerman E. 1987. 'Aspects of Transferability in Second Language Acquisition'. Ph.D Thesis. University of Nijmegen.

Kellerman E. and M. Sharwood (eds.). 1986. Crosslinguistic Influence in Second Language Acquisition. Oxford : Pergamon Press.

Khaldi K. 1981. 'The Use of the Strategy of Transfer by Learners with a Bilingual Background'. M.Litt Thesis. University of Edinburgh.

Laufer B. 1988. 'The concept of synforms (similar lexical forms) in vocabulary acquisition'. Language and Education 2/2 : 113-32.

Laufer B. 1989. 'A factor of difficulty in vocabulary learning : Deceptive transparency'. AILA Review 6 : 10-20.

Lightbown P.M. and G. Libben. 1984 'The recognition and use of cognates by L2 learners' in R.W. Andersen (ed.) 1984 : 393-418.

Lyons J. 1981. Language, Meaning and Context. London : Fontana.

Meisel, J. 1983. 'Transfer as a second language strategy'. Language and Communication 3 : 11-46.

Nababan P.W.J. 1971. 'A note on transfer and interference in foreign language learning'. Working Papers in Linguistics 3/4 : 17-24. Department of Linguistics, University of Hawaii.

Ringbom H. 1978. The influence of the mother tongue on the translation of lexical items'. Interlanguage Studies Bulletin 3/1 : 80-101.

Ringbom H. 1983. 'Borrowing and lexical transfer'. Applied Linguistics 4(3) : 207-12.

Ringbom H. 1986. 'Crosslinguistic influence and the foreign language learning process' in E. Kellerman and M. Sharwood (eds.) 1986 : 150-162.

Ringbom H. 1987. The Role of the First Language in Foreign Language Learning. Clevedon : Multilingual Matters.

Rivers W.M. 1979. 'Learning a sixth language : an adult learner's daily dairy'. Canadian Modern Language Review 36/1 : 67-82.

Savard J.G. and L. Laforge (eds). 1981. Proceedings of the 5th Conference of l'Association Internationale de Linguistique Appliquée. Quebec : Les Presses de l'Université Laval.

Schachter J., A.F. Tyson and F.J. Diffley. 1986. 'Learner intuitions of grammaticality. Language Learning 26/1 : 67-76.

Singh R. and S. Carroll. 1979. 'L1, L2, and L3'. Indian Journal of Applied Linguistics 5/1 : 51-63.

Sweet H. 1964. The Practical Study of Languages. London : Oxford University Press.

Tenjoh-Okwen T. 1985. 'An Analysis of Some Aspects of the Interlanguage of Francophone Cameroonian Learners of English. Ph.D Thesis. Department of English : University of Ibadan.

Thomas J. 1988. 'The role played by metalinguistic awareness in second and third language learning'. Journal of Multilingual and Multicultural Development 9/3 : 235-46.

Uljin J.M., S. Wolfe and A. Donn. 1981. 'The Lexical Transfer Effect of French Knowledge in the Acquisition of English by Native Vietnamese Speakers'. Report No.6 F.L.A. Research THE, Eindhoven University of Technology.

Vildomec V. 1963. Multilingualism. Leyden : Sythoff.

Wickstrom Y. 1980. 'Mother Tongue Interference and Third Language Interference with Special Reference to Finns Learning Swedish and English'. M.A Dissertation. Abo Akademi : Department of English.

White L. 1987. 'Markedness and second language acquisition : the question of transfer'. Studies in Second Language Acquisition 9 : 261-85.

Appendices

A. Sentence Completion Task (Version 1)

Complete the following sentences with the missing word. The information supplied in each sentence will help to choose the appropriate word. Only one word answers should be given and you should not give any word already used in the sentence as your answer. Do not hesitate to ask me if there is a word used in the sentences that you do not understand.

1 There must have been a .... in their friendship because I have not seen them together for ten months. The problem is that they do not want to tell anyone the truth.

2 Two Americans were deported from Iraq after it was found out that they worked for the CIA but they insisted that they were not involved in ....

(The remaining items were exactly as in Version 2 below, without the French prompts.)

B. Sentence Completion Task (Version 2)

Complete the following sentences with the appropriate missing word. The French translation of the missing word has been provided to help you. **Only one word answers** should be given. Do not hesitate to ask me if there is a word used in the sentences that you do not understand.

1 There must have been a .... in their friendship because I have not seen them together for ten months. The problem is that they do not want to tell anyone the truth (rupture).

2 Two Americans were deported from Iraq after it was found out that they worked for the CIA but they insisted that they were not involved in.............................. (espionnage)

3 He loves food so much that everyone is amazed at his .... Even his own children have to keep away from him while he is eating (gourmandise).

4 He .... his work to eat lunch (interrompre)

5 He can now build a house because he has got a £100,000 bank ....(crédit)

6 What is the best and shortest .... from here to Switzerland? (route)

7 Two bombs exploded shortly before the .... of the cabinet meeting while the ministers were still waiting for the Prime Minister (commencement).

8 Prince Rwagasore was .... by the enemies of Uprona (assassiné).

9 Kuwait was .... by the Allies after seven months of occupation (libéré).

10 He performed a lot of .... in the laboratory (expériences).

11 He .... all his money in the bank and forgot to keep some for the weekend shopping (déposer).

12 All the athletes were suffering from .... at the end of the marathon race. They were all exhausted (fatigue).

13 Our landlady asked for a £50 .... to cover any damage we might cause during our stay (caution).

14 Our University does not have enough .... because very few students are attracted to the Department of Physics. But it has a lot of mathematicians (physiciens).

15 His mind became .... as a result of his long imprisonment (dérangé).

16 The criminal's statements .... a local politician in the crime (impliquer).

17 The store sells .... from all over the world. It sells very few items which are local products (marchandises).

18 You promised to .... all the money I paid for your clothes and you cannot change your mind now (rembourser).

19 He was so .... asleep that he completely could not hear the fire alarm (profondément).

20 You should not buy this .... because it has already been involved in several accidents. Besides,you said you would prefer a Ford to a Peugeot (véhicule).

21 She was found guilty of .... because she did not look after her children properly. Her irresponsibility was condemned by many parents (négligence).

22 In Britain the academic year .... in October just like in Burundi (commencer).

23 The thieves .... the car they had stolen on the road and ran away while the police were following them (abandonner).

24 Inter Star is the most experienced football .... in Burundi (formation).

25 The cabinet meeting will .... at five o'clock. So the Prime Minister will not be available until that time (se terminer).

26 The Iraqi troops .... on the 43rd day of the Gulf War, which was the day the war ended (capituler).

27 The chairman of the conference forgot to .... the speakers to the audience (présenter).

28 Please do not .... this seat because it has been reserved (occuper).

29 While we were eating lunch, my brother asked me to .... him the salt (passer).

30 He tried hard to ....his theory to the experts who attended his lecture (expliquer).

C. Lexico-Semantic Acceptability Judgment

Using a scale of 5 points, indicate the degree to which you accept the underlined words as appropriately used. Along the scale point 5 means completely acceptable, 1 means completely unacceptable and 4, 3, and 2 are intermediate points. Give your answer by putting a **cross** in **only one** of the five boxes.

1 My brother is a veterinary. He is a doctor for animals.
5[ ]4[ ]3[ ]2[ ]1[ ]

2 The Allies demanded that Iraq accept all the 12 UN resolutions.
5[ ]4[ ]3[ ]2[ ]1[ ]

3 If you ignore my advice, you will regret it later on.
5[ ]4[ ]3[ ]2[ ]1[ ]

4 The Finance Minister remarked that the country's economy was in recession.
5[ ]4[ ]3[ ]2[ ]1[ ]

5 He has just attained the age of twenty.
5[ ]4[ ]3[ ]2[ ]1[ ]

6 His fiancée attended him all through his illness.
5[ ]4[ ]3[ ]2[ ]1[ ]

7 The termination of hostilities in the Gulf War was awaited by many people all over the world.
5[ ]4[ ]3[ ]2[ ]1[ ]

8 The army commander was devastated by the news that 50 of his soldiers had been killed by friendly fire.
5[ ]4[ ]3[ ]2[ ]1[ ]

9 Mr Major succeeded Mrs Thatcher as the Prime Minister of the U.K.
5[ ]4[ ]3[ ]2[ ]1[ ]

10 The referee cautioned the player three times before he sent him off.
5[ ]4[ ]3[ ]2[ ]1[ ]

11 Who administrates your financial affairs?
5[ ]4[ ]3[ ]2[ ]1[ ]

12 Prime Minister Major saluted the courage and conduct of the British troops during the Gulf War.
5[ ]4[ ]3[ ]2[ ]1[ ]

13 President Bush commended the US forces for their brilliant victory.
5[ ]4[ ]3[ ]2[ ]1[ ]

14 If you got too fatigued, your heart would get worse.
5[ ]4[ ]3[ ]2[ ]1[ ]

15 Which ground is the venue for the next football match?
5[ ]4[ ]3[ ]2[ ]1[ ]

16 The club members have nominated a new president.
5[ ]4[ ]3[ ]2[ ]1[ ]

17 She was very sympathetic when I failed my exam.
5[ ]4[ ]3[ ]2[ ]1[ ]

18 His father gave him a serious reprimand for damaging his car.
5[ ]4[ ]3[ ]2[ ]1[ ]

19 You will not ameliorate the situation by giving a long explanation.
5[ ]4[ ]3[ ]2[ ]1[ ]

20 His behaviour is inexcusable.
5[ ]4[ ]3[ ]2[ ]1[ ]

21 He received a large sum of money as recompense for stealing the enemy's war plan.
5[ ]4[ ]3[ ]2[ ]1[ ]

22 In many countries, leaders are overthrown by their own entourage.
5[ ]4[ ]3[ ]2[ ]1[ ]

23 The patient died in the theatre while he was being operated upon.
5[ ]4[ ]3[ ]2[ ]1[ ]

24 The British Government expressed its grave concern about the treatment of POWs (Prisoners Of War) by the Iraqi Government.
5[ ]4[ ]3[ ]2[ ]1[ ]

25 The company made large interests from exports.
5[ ]4[ ]3[ ]2[ ]1[ ]

26 Children have a lot more liberty now than they used to.
5[ ]4[ ]3[ ]2[ ]1[ ]

27 The siege for the United Nations is in New York.
5[ ]4[ ]3[ ]2[ ]1[ ]

28 The song will undoubtedly become a massive hit.
5[ ]4[ ]3[ ]2[ ]1[ ]

29 The situation necessitated his immediate return.
5[ ]4[ ]3[ ]2[ ]1[ ]

30 I can aid you in your research by providing you with some data.
5[ ]4[ ]3[ ]2[ ]1[ ]
31 British people still estimate Mrs Thatcher as an outstanding politician.
5[ ]4[ ]3[ ]2[ ]1[ ]
32 She has bought a nice 1992 agenda.
5[ ]4[ ]3[ ]2[ ]1[ ]
33 The Iraqi army should depose Saddam Hussein for the good of the country.
5[ ]4[ ]3[ ]2[ ]1[ ]
34 She promenaded her children through the park.
5[ ]4[ ]3[ ]2[ ]1[ ]
35 It is difficult to comprehend the behaviour of that man.
5[ ]4[ ]3[ ]2[ ]1[ ]
36 I am persuaded that multiparty systems do not necessarily mean democracy.
5[ ]4[ ]3[ ]2[ ]1[ ]
37 Much of the government's revenue comes from exports.
5[ ]4[ ]3[ ]2[ ]1[ ]
38 He is such an authoritative father that no child can object to his decisions.
5[ ]4[ ]3[ ]2[ ]1[ ]
39 His lecture comprehended several aspects of the topic.
5[ ]4[ ]3[ ]2[ ]1[ ]
40 Although this medicine does not cure the illness, it alleges the pain.
5[ ]4[ ]3[ ]2[ ]1[ ]
41 The assassin of Gandhi is still unknown.
5[ ]4[ ]3[ ]2[ ]1[ ]
42 The President pardoned all the political prisoners.
5[ ]4[ ]3[ ]2[ ]1[ ]
43 The function was attended by many dignitaries.
5[ ]4[ ]3[ ]2[ ]1[ ]
44 Multiparty system is an important actuality in African politics today.
5[ ]4[ ]3[ ]2[ ]1[ ]
45 Large lorries are a menace on narrow roads.
5[ ]4[ ]3[ ]2[ ]1[ ]
46 He became the child's guardian when her parents were killed in a car crash.
5[ ]4[ ]3[ ]2[ ]1[ ]
47 The customs officer was found guilty of concussion.
5[ ]4[ ]3[ ]2[ ]1[ ]
48 Iraqi demonstrators were chanting slogans against President Bush.
5[ ]4[ ]3[ ]2[ ]1[ ]
49 The doctor administered the drugs to that patient.
5[ ]4[ ]3[ ]2[ ]1[ ]
50 The problem he is faced with is formidable.
5[ ]4[ ]3[ ]2[ ]1[ ]

THE UNIVERSITY OF EDINBURGH

Institute for Applied Language Studies (IALS)

The Institute's English courses are designed to meet specialist language learning needs and run all year round. Specialist courses are available for law, business and management, medicine and also academic preparation.

As well as English courses, IALS is a centre for the teaching of other languages; R & D; distance learning; consultancy and courses abroad; publications and materials for language learning; worldwide language development projects such as the Edinburgh Project in Extensive Reading.

IALS
a creative language learning environment

Institute for Applied Language Studies,
21 Hill Place, Edinburgh, EH8 9DP

Tel: 031-650 6200 Telex: 727442 UNIVED G Fax: 031-667 5927

TEACHERS' COURSES AT IALS

The Institute for Applied Language Studies offers a wide range of courses and other services for teachers of English and other languages.

Certificated courses (full-time)

- UCLES/RSA Diploma in the Teaching of English as a Foreign Language to Adults (one term, commencing October and March)
- Advanced Certificate in English Language Teaching (one term, commencing October and March)
- Advanced Certificate in English Studies (two terms, commencing October)
- Higher Advanced Certificate in English Studies (three terms, commencing October)

Certificated courses (distance-learning)

- Advanced Certificate in English Language Teaching (one year, commencing June for individual enrolments)
- RSA Diploma in the Teaching of Foreign Languages (one year, commencing September)

Short summer courses

- Teaching and Learning English
- Teaching English for Specific Purposes
- Teaching English for Medical Purposes
- Teaching English for Business Purposes
- Grammar and the Communicative Teaching of English
- Drama for TEFL
- CEELT Preparation course
- English for EFL/Applied Linguistics

Trainer Training

- A modular course, available only to sponsored groups

Tailor-made versions of these courses, together with seminars on specialised topics and a language-teaching consultancy service, can also be supplied overseas.

For further information write to "Teachers' Courses",
IALS, University of Edinburgh, 21 Hill Place, Edinburgh EH8 9DP.
Tel. 031 650 6200   Fax. 031 667 5927

Contents of previous issues: EWPAL 1 (1990)

FALLING FROM GRICE: THE IDEOLOGICAL TRAP OF PRAGMATIC STYLISTICS
Peter Tan

UG, SLA AND LANGUAGE TEACHING: FUTURE OF AN ILLUSION?
Liam Rodger

WHAT, IF ANYTHING, IS ENGLISH FOR LITERARY STUDIES?
Brian Parkinson

TRAINER TRAINING, SIMULATION AND THE FEEDBACK IMPERATIVE
Ian McGrath and Sevgi Altay

THE WASHBACK EFFECT OF A TEXTBOOK-BASED TEST
T.R. Khaniya

TALKS: WHAT NEGOTIATORS THINK DRIVES A HARD BARGAIN
Mary Ann Julian

THE EFFECT OF CO-OPERATIVE TEACHING ON THE QUANTITY AND QUALITY OF TEACHER FOREIGN LANGAUGE DISCOURSE IN THE FRENCH CLASSROOM
Carole E.M. Franklin

DO THE SIMIALRITIES BETWEEN L1 AND L2 WRITING PROCESSES CONCEAL IMPORTANT DIFFERENCES?
Ana Frankenberg-Garcia

COURSE DEVELOPMENT IN NATURALISTIC PERSPECTIVE
Gibson Ferguson and Gillian D. Brown

HAVE THE GOAL POSTS BEEN MOVED? A consideration of goals of African speakers of English in the context of the emergence of non-native varieties of English
Esther Daborn

SER ou NAO SER? A study of Spanish-Portuguese cross linguistic influence
Cathy Benson

Contents of previous issues: EWPAL 2 (1991)

DISCURSIVE ASPECTS OF METAFICTION: A NEO-ORAL AURA?
Alexandra Georgakopoulou

THE REACTION OF LEARNERS TO TAPE-BASED LISTENING COMPREHENSION MATERIALS IN FRENCH, SPANISH AND ITALIAN 'COMMUNITY CLASSES'
Brian Parkinson, Giulia Dawson, Lucila Makin, Hélène Mulphin

INTERLANGUAGE LEXIS: AN INVESTIGATION OF VERB CHOICE
David J. Hill

AN INVESTIGATION OF A TIMETABLED SELF-ACCESS SESSION IN A GENERAL ENGLISH PROGRAMME
Sheena Davies, Eileen Dwyer, Anne Heller, Kate Lawrence

IS IT OR IS IT NOT INTERLANGUAGE? A HEAD-ON CONFRONTATION WITH NON-NATIVE ENGLISH
L.K. Owusu-Ansah

CONDITIONALS AND THE EXPRESSION OF REGRET AND RELIEF: TOWARDS A FRAGMENT FOR A COMMUNICATIVE GRAMMAR
Gibson Ferguson

ARGUMENTS FOR AND AGAINST FREE VARIATION
S.B. Makoni

BIBLIOGRAPHIC PRESENTATION
Tony Lynch and Ian McGrath

CROSSLINGUISTIC INFLUENCE IN A BILINGUAL CLASSROOM; THE EXAMPLE OF MALTESE AND ENGLISH
Antoinette Camilleri

ASSESSING THE READABILITY OF MEDICAL JOURNAL ARTICLES: AN ANALYSIS OF TEACHER JUDGEMENTS
Gibson Ferguson and Joan Maclean

LITERARY DISCOURSE AND IRONY: SECRET COMMUNION AND THE PACT OF RECIPROCITY
Sonia S'hiri

Contents of previous issues: EWPAL 3 (1992)

ASYMMETRIC RESETTING OF THE NON-EMPTY TOPIC PARAMETER BY CHINESE-SPEAKING LEARNERS OF ENGLISH
Boping Yuan

THE SOCIOLINGUISTIC STATUS OF ENGLISH IN MALTA
Antoinette Camilleri

INTERLANGUAGE PHONOLOGY: THE PERCEPTUAL DEVELOPMENT OF DURATIONAL CONTRASTS BY ENGLISH-SPEAKING LEARNERS OF JAPANESE
Kayoko Enomoto

PARTICIPANT REFERENCE IN GREEK PERSONAL EXPERIENCE NARRATIVES
Alexandra Georgakopoulou

READING, CULTURE AND COGNITION
Martin Gill

CLUSTER ANALYSIS AND THE INTERLANGUAGE LEXICON
David J. Hill

INTERACTIONAL LISTENING TASKS: A COMPARATIVE STUDY OF STRATEGY AND PRACTICE TEACHING APPROACHES
Vanessa Luk

LESSON BEGINNINGS
Ian McGrath, Sheena Davies and Hélène Mulphin

WHERE TRUE POWER LIES: MODALITY AS AN INDICATION OF POWER IN TWO INSTITUTIONALISED DOMAINS OF LANGUAGE USE
L.K. Owusu-Ansah

TENSE, ASPECT AND THE BUSCONDUCTOR HINES - THE LITERARY FUNCTION OF NON-STANDARD LANGUAGE IN THE FICTION OF JAMES KELMAN
Liam Rodger

A PRAGMATICS OF VERBAL IRONY IN LITERARY DISCOURSE: AN EXAMPLE FROM DRAMA
Sonia S'hiri

AN INVESTIGATION INTO LEARNERS' REACTIONS TO LEARNER TRAINING FOR SELF-ACCESS LISTENING
Elspeth Wardrop and Kenneth Anderson

EWPAL 4 contributors' addresses:

Joan Cutting
Alan Davies
Ardeshir Geranpayeh
Matutin Sikogukira
Behzad Ghonsooly

Dept. of Applied Linguistics
14 Buccleuch Place
Edinburgh EH8 9LN
Scotland - UK

Gibson Ferguson
Aileen Irvine
Tony Lynch
Brian Parkinson

IALS
21 Hill Place
Edinburgh EH8 9DP
Scotland - UK

Martin Gill

Freie Universität Berlin
ZE Sprachlabor
Habelschwerdter Allee 45
1000 Berlin 33

SUBSCRIPTION / EXCHANGE FORM FOR EWPAL

Name/Institution

Address

Please tick the relevant items

Issue number *	EWPAL number 2 (1991)	
	EWPAL number 3 (1992)	
	EWPAL number 4 (1993)	
	EWPAL number 5 (1994)	
Type of subscription (price per issue)	individual surface / EC	£5
	individual non-EC air mail	£7
	institutional surface / EC	£6
	institutional non-EC air mail	£8
	institutional exchange	

(For exchanges) I enclose a current copy of our institutional publication entitled

....

Total amount of payment: £ ....

Method of payment

Sterling cheque (enclosed) to University of Edinburgh

Sterling transfer (reference no ....)
to the Royal Bank of Scotland, Nicolson Street, Edinburgh,
Scotland - bank code 83-19-19, account IALS 00181202

Please send this form completed to

EWPAL subscriptions
IALS
21 Hill Place
Edinburgh EH8 9DP
Scotland - UK

* EWPAL number 1 (1990) is now out of print, but copies of single articles are available on request, free of charge, to subscribers to any of the subsequent issues.

ISSN 0959-2253

BEST COPY AVAILABL